AIX-MARSEILLE UNIVERSITE
FACULTE DE MÉDECINE DE MARSEILLE
ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

UNIVERSITE CHEIKH ANTA DIOP
ECOLE DOCTORALE DES SCIENCES DE LA VIE, DE LA SANTE ET DE L'ENVIRONNEMENT
FACULTE DES SCIENCES ET TECHNIQUES

THÈSE DE DOCTORAT

EN CO-TUTELLE

Spécialités: Maladies Transmissibles et Pathologies Tropicales (AMU) / Parasitologie (UCAD)

Présentée par

Monsieur **Cheikh Ibrahima LO**

Né le 10 mai 1985 à Keur Khar Kane, Gossas (Sénégal)

REPERTOIRE DES BACTERIES IDENTIFIEES PAR MALDI-TOF EN AFRIQUE DE L'OUEST

Soutenue publiquement
A LA FACULTÉ DE MÉDECINE DE MARSEILLE
Le 7 Décembre 2015
Devant le Jury composé de:

Président :	M. Pierre Edouard Fournier	Professeur, AMU/France
Membre du Jury :	M. Bhen Sikina Toguebaye	Professeur, UCAD/Sénégal
Rapporteur :	Mme. Marie Kempf	Docteur CHU-Angers/France
Rapporteur :	Mme. Amy Gassama Sow	Professeur UCAD/Sénégal
Directeur de Thèse:	Mme Florence Fenollar,	Professeur, AMU/France
Co-directeur de Thèse:	M. Ngor Faye	Maître de Conférences, UCAD/Sénégal

LABORATOIRE D'ACCUEIL

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63, CNRS 7278, IRD 198, Inserm 1095; 13005 Marseille, France.

Remerciements

Tout d'abord, je remercie Madame **Florence FENOLLAR**, Professeur à Aix-Marseille Université. En tant que directrice principale de ma thèse, elle m'a guidé dans mon travail et m'a aidé à trouver des solutions pour avancer. Elle a consacré un temps énorme pour moi et n'a ménagé aucun effort pour l'atteinte des objectifs de cette thèse. J'ai appris le franc parlé, la rigueur, l'amour de la recherche et tant d'autres qualités qui me serviront certes à l'avenir. Je vous adresse ma profonde gratitude.

Ensuite au Professeur **Ngor FAYE** de l'Université Cheikh Anta Diop qui a toujours était à nos côtés depuis les premières années à l'UCAD et qui a guidé nos pas au chemin de la recherche. Votre aide et vos conseils m'ont été bénéfiques. Qu'il nous soit permis de vous témoigner ici, toute ma reconnaissance.

Merci encore au Professeur **Pierre-Edouard FOURNIER** d'Aix-Marseille Université, qui nous a fait l'honneur d'être le Président de jury de cette thèse avec preuve d'une disponibilité réelle qui n'est qu'une de vos qualités humaines. Merci infiniment.

Je remercie également le Professeur **Bhen Sikina TOGUEBAYE** de l'UCAD, qui nous a fait l'honneur de participer au jury de cette thèse et d'accepter de juger ce travail.

Il me paraît important de remercier le Docteur **Marie KEMPF**, Maître de Conférences à la Faculté de Médecine d'Angers, et le Professeur **Amy GASSAMA SOW**, chercheuse à l'Institut Pasteur de Dakar pour avoir bien voulu juger mon modeste travail de thèse. En acceptant d'être rapporteur de mon jury de thèse vous me faites un grand honneur. Je vous adresse ici ma profonde gratitude.

Je remercie solennellement le Professeur **Didier RAOULT** de m'avoir accueilli dans son laboratoire et d'avoir financé mes travaux de recherche par le biais de la **Fondation Infectiopôle Sud** qu'il dirige.

Je ne remercierai jamais assez des amis comme **Masse Sambou, Alpha Kabinet Keita, Jérôme Terras, Valérie Filosa, Niokhor Dione, Sylvain Buffet, Micheline Pitaccolo, Oleg Mediannikov, Bécaye Fall, Bissoume Samb-Ba, Roshan Padmanabhan, Constentin Diémé, Carine Couderc, Catherine Robert, Senthil Alias Sankar, Seydina Mohammed Diene, Aurélie Caputo, Hubert Bassène, Annick Abeille**... ainsi que tous mes collègues, les techniciens et le personnel administratif de l'URMITE qui ont tous d'une manière ou d'une autre participer à la réussite de ma thèse.

Je remercie également mes parents **Cheikh Abdoulaye LO** et **Sokhna Mame Faty Ndiaye** pour leur assistance et leurs prières perpétuelles. Merci également à **mes frères et sœurs** et **à celle** qui a fait, pour moi, de la patience son arme…

Dieureudieuf Borom Touba
Diarama Cheikh Ibrahima Fall

Sommaire

Avant-Propos

Le format de présentation de cette thèse correspond à une recommandation de la spécialité Maladies Infectieuses et Microbiologie, à l'intérieur du Master des Sciences de la Vie et de la Santé qui dépend de l'Ecole Doctorale des Sciences de la Vie de Marseille.

Le candidat est amené à respecter des règles qui lui sont imposées et qui comportent un format de thèse utilisé dans le Nord de l'Europe et qui permet un meilleur rangement que les thèses traditionnelles. Par ailleurs, la partie introduction et bibliographie est remplacée par une revue envoyée dans un journal afin de permettre une évaluation extérieure de la qualité de la revue et de permettre à l'étudiant de commencer le plus tôt possible une bibliographie exhaustive sur le domaine de cette thèse.

Par ailleurs, la thèse est présentée sur article publié, accepté ou soumis associé d'un bref commentaire donnant le sens général du travail. Cette forme de présentation a paru plus en adéquation avec les exigences de la compétition internationale et permet de se concentrer sur des travaux qui bénéficieront d'une diffusion internationale.

Professeur Didier RAOULT

Résumé

Le répertoire des bactéries est mal connu en Afrique du fait des méthodes d'étude essentiellement basées sur des techniques de culture sur milieux simples associées à des tests biochimiques, ce qui ne permet pas son exploration. Il a néanmoins été récemment bouleversé par l'usage systématique de la spectrométrie de masse de type MALDI-TOF MS.

Au cours de nos travaux de thèse de doctorat, nous avons utilisé deux types de spectromètres de masse: le MALDI-TOF Vitek MS, nouvellement installé à Dakar; le MALDI-TOF Microflex LT, installé à Marseille. Au Sénégal, les souches étudiées ont surtout été isolées à l'Hôpital Principal de Dakar (HPD) et, ou parfois, dans les dispensaires partenaires de l'Institut de Recherche pour le Développement (IRD). Après culture sur milieux gélosés, elles sont d'abord testées avec le MALDI-TOF Vitek puis envoyées à Marseille où l'identification est confirmée avec le MALDI-TOF Microflex LT, ou complétée par l'amplification couplée au séquençage du gène de l'ARN 16S ribosomique et/ou des gènes de ménages comme le *rpoB*. Quant à l'ADN total des génomes des nouvelles espèces bactériennes, il est extrait et entièrement séquencé par des techniques de biologie moléculaire.

Les résultats que nous avons obtenus ont montré, pour la première fois, que la technique du MALDI-TOF est efficace et tout à fait adaptée en Afrique pour le diagnostic spécifique de routine. Cette performance a conduit le laboratoire clinique de l'HPD à opter pour son utilisation à la

place des traditionnelles techniques d'identification phénotypique telles que les galeries API. Nous avons également confirmé que le MALDI-TOF est un puissant outil d'identification des espèces bactériennes rarement impliquées dans les maladies infectieuses humaines. De plus, cet outil nous a permis de détecter sept nouvelles espèces de bactéries isolées pour la première fois chez l'homme. En Afrique, il faudrait donc multiplier l'installation de spectromètre de masse MALDI-TOF, ou mettre en place des réseaux autour de plateformes MALDI-TOF sous-régionales partagées entre plusieurs structures sanitaires et/ou de recherche.

Abstract

The Africa bacteria repertory is unfamiliar because the available tools in this region are not allowed its best knowledge. In fact, bacteria are most often identified using culture techniques on simple media and biochemical tests which enable the identification of some common characters. These methods do not facilitate an exhaustive knowledge of the bacterial repertory; consequently they have recently been revolutionized by the systematic use of MALDI-TOF mass spectrometry (MS).

In our thesis we used two mass spectrometers, respectively, MALDI-TOF Vitek MS currently installed at Dakar (Senegal) and MALDI-TOF Microflex LT installed in Marseille (France). The strains are isolated mainly in the Hôpital principal de Dakar or sometimes in IRD's dispensaries located in Senegal. After growth on agar media, strains were firstly tested with the MALDI-TOF Vitek then shipped to Marseille for two reasons namely: to confirm their identification with the MALDI-TOF Microflex and to complete a misidentification from Dakar by amplifying and sequencing the 16S RNA gene or a housekeeping gene like *rpoB* gene. Lastly the whole genome of bacterial new species will be extracted and completely sequenced by bio-molecular methods of new generation. Our research showed for the first time that the MALDI-TOF technology is effective and quite suitable for the identification routine of bacterial species in Africa. Thus the performance of MALDI-TOF for high throughput identifying of

microorganisms induced the clinical laboratory of the HPD to drop conventional phenotypic techniques of identification such as API galleries systems. In addition we have also confirmed that MALDI-TOF is a powerful tool for identifying bacterial species rarely involved in human infectious diseases. Thus in adopting the MALDI-TOF as a first-line tool in bacterial identification before Gram staining or other techniques of phenotypic identifications based on chemical characteristics, we discovered seven new species of bacteria isolated for first time in humans. Microbial identification using MALDI-TOF MS is currently feasible in Africa. Its performance and effectiveness in routine diagnosis of clinical microbiology laboratories have been proven. It is necessary either to increase the installation of MALDI-TOF, or establishing a network around a shared MALDI-TOF platform between several structures located in the same area, especially in the underdeveloped countries of Africa amortization of investment costs of the device, because it allowed reducing the time of reporting results and indirectly facilitating better care for patients.

Introduction

L'utilisation de la spectrométrie de masse de type MALDI-TOF (Matrix Assisted Laser Desorption Ionization-Time Of Flight) pour l'identification des bactéries a beaucoup révolutionné la pratique en microbiologie clinique au cours de ces dernières années (1,2). En effet, le MALDI-TOF est un appareil (spectromètre de masse) qui permet d'identifier des microorganismes par analyse de leurs composantes protéiques selon leur rapport masse/charge (3); cette technique est efficace et offre surtout un gain de temps et d'argent par rapport aux anciennes méthodes utilisées (4). Depuis ces dix dernières années, la technologie de la spectrométrie de masse MALDI-TOF a été développée dans les laboratoires de diagnostic pour devenir une méthode de routine pour l'identification des microorganismes à partir de prélèvements cliniques (5,6). En effet, la spectrométrie de masse MALDI-TOF est une technique bien établie qui permet une identification à haut débit, précise et reproductible de microorganismes d'intérêt et ne demande pas une préparation fastidieuse et une connaissance préalable de l'échantillon (6–10).

Le spectromètre de masse MALDI-TOF est un instrument qui a subi une évolution progressive dans le domaine de l'identification des microorganismes. En 1975, des scientifiques ont proposé une méthode associant la spectrométrie de masse à la pyrolyse pour la détection de protéines bactériennes (3). Puis la technique d'ionisation électrospray a été rapportée par Masamichi Yamashita et John Fenn en 1984 (11). En outre, le MALDI-TOF MS a été développé dans les années 80 par Karas

et Hillenkamp en Allemagne et Tanaka *et al*. au Japon (12,13). En effet, en 1985, Karas *et al*. découvrent une méthode plus sensible appelée «matrix assisted laser desorption» (MALDI) pour la première fois en utilisant un petit composé organique comme matrice (14). Plus tard en 1987, on assiste à une première description d'un instrument TOF-TOF (Time-of-flight) rapportée par Schey *et al*.(15). En 1988, le premier appareil MALDI-TOF MS est commercialisé par la société Shimadzu. Ensuite le prix Nobel de chimie a été décerné à Koichi Tanaka et Michael Karas en 2002 pour leurs travaux visant à développer des méthodes douces appelées « soft desorption ionization » (16). Enfin, en 2009 l'équipe du Professeur Raoult a démontré les performances et la rentabilité du MALDI-TOF dans l'identification systématique des bactéries et des champignons pour le diagnostic de routine (2). Ainsi, on assiste à une explosion des travaux effectués avec le MALDI-TOF à travers le monde de la microbiologie clinique (1,4,17,18). Le MALDI-TOF que nous utilisons comme tout spectromètre de masse est composé de quatre éléments essentiels qui sont : une source d'ionisation qui provoque le passage de l'échantillon en phase gazeuse et l'ionisation des molécules, un analyseur qui sépare les ions selon leur rapport masse sur charge (m/z), un détecteur permettant de transformer le courant ionique en courant électrique et enfin un système informatique traitant le signal pour analyser et afficher les spectres correspondant à l'échantillon étudié (19,20).

Les bactéries sont des microorganismes vivants microscopiques très nombreux (23). On les retrouve dans presque tous les biotopes du monde. Leur nombre est estimé à quatre à six quintillions ($4x10^{30}$ à

$6x10^{30}$) (24) dont une minorité est pathogène pour l'homme. L'identification des microorganismes est une étape décisive pour la prise en charge des patients en vue d'élaborer une antibiothérapie efficace. Pour identifier une bactérie, les laboratoires de microbiologie cliniques ont recours à diverses tests phénotypiques reposant sur l'analyse de la morphologie après la coloration de Gram, les caractéristiques de culture et de croissance ainsi que les caractéristiques biochimiques (25). Même si certains tests se font rapidement, l'identification finale du microorganisme peut durer plusieurs heures voire plusieurs jours dans le cas des bactéries fastidieuses (26).

L'identification par biologie moléculaire notamment l'amplification et le séquençage du gène de l'ARN 16S ribosomique a permis la description de nouvelles espèces de bactéries (27,28). Cependant ces techniques moléculaires malgré leur efficacité et leur précision, ne sont pas utilisées pour l'identification en routine des pathogènes, du fait de leur coût élevé, de leurs produits exigeants une conservation délicate et de la nécessité de disposer du matériel et de personnel spécifiquement formé.

Par ailleurs, l'utilisation de la spectrométrie de masse de type MALDI-TOF dans les laboratoires de microbiologie clinique, qui permet une identification précise, rapide et à haut débit des microorganismes, offre une nouvelle approche dans la lutte contre les maladies infectieuses et émergentes qui sévissent en Afrique et en zone tropicale.

L'objectif général de notre thèse de doctorat est d'évaluer l'intérêt de l'utilisation et la faisabilité de la technologie de la spectrométrie de masse de type MALDI-TOF dans le répertoire des bactéries en Afrique. Plus spécifiquement, il s'agit: (1) d'identifier de nouvelles espèces de bactéries; (2) d'incrémenter la base de données MALDI-TOF; (3) de faire la caractérisation des génomes bactériens etc. Notre mémoire de thèse de doctorat s'articule sur 3 chapitres.

Dans le premier chapitre, nous rapportons une revue de la littérature qui porte sur une synthèse des données actuelles permettant de discuter de l'intérêt de l'utilisation du spectromètre de masse MALDI-TOF dans les laboratoires de microbiologie clinique en Afrique et en zone tropicale.

Dans le second chapitre, nous présentons nos travaux de recherches effectués sur la spectrométrie de masse MALDI-TOF et complétés en partie par la biologie moléculaire. Ces travaux ont générés deux articles scientifiques détaillés dans le chapitre 2. Ces deux articles confirment la performance et la faisabilité de l'identification par spectromètre de masse MALDI-TOF des bactéries isolées dans les prélèvements cliniques de patients hospitalisés à l'Hôpital Principale de Dakar.

Enfin, dans le chapitre 3, nous décrivons 7 des nouvelles espèces bactériennes isolées et identifiées durant notre Thèse. Nous rapportons notamment leurs caractéristiques phénotypiques et génotypiques. Nous y présentons aussi la description d'une bactérie *Pantoea septica*, pour laquelle aucun génome n'avait été décrit jusqu'à présent. Chacune de ces 8 bactéries a donné lieu à une publication scientifique.

Chapitre 1 :

Revue de la littérature sur la valeur du MALDI-TOF en microbiologie clinique et dans la lutte contre les maladies infectieuses en Afrique.

Article 1:

« Value of MALDI-TOF MS in Clinical Microbiology and Infectious Diseases in Africa and tropical areas ».

Cheikh Ibrahima Lo[1], Bécaye Fall[2], Bissoume Sambe-Ba[2], Ngor Faye[3], Christophe Flaudrops[1], Boubacar Wade[2], Didier Raoult[1], Florence Fenollar[1*]

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de Médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France Tel: +33 491 385 517, Fax: +33 491 387 772

[2]Hôpital Principal de Dakar, Laboratoire de Microbiologie, Dakar, Senegal

[3]Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann, Senegal

Soumis à:

The American Journal of Tropical Medicine and Hygiene

Résumé de la revue : ***« Value of MALDI-TOF MS in Clinical Microbiology and Infectious Diseases in Africa and tropical area s »***

L'utilisation de la spectrométrie de masse MALDI-TOF dans le cadre de l'identification des microorganismes a apporté un grand tournant dans la microbiologie clinique en 2009. Actuellement, un tel outil est présent dans presque tous les continents malgré son coût encore assez élevé. En effet, il est aujourd'hui possible de détecter par MALDI-TOF des microorganismes directement contenus dans des prélèvements liquides tels que le sang après pré-incubation. Actuellement il y a un essor fulgurant de l'utilisation de la spectrométrie de masse visant à explorer plusieurs chemins permettant de réduire le délai de rendu des résultats de diagnostic et de remplacer certaines méthodes conventionnelles utilisées dans certains laboratoires. Par ailleurs pour identifier une bactérie par le MALDI-TOF, nous suivons en général les étapes suivantes:

a- Prélever à l'aide d'une anse une colonie bactérienne pure (21,22)
b- Déposer finement la colonie sur un à deux puits de la plaque métallique
c- Ajouter une goutte de matrice (α-cyano-4-hydroxy-cinnamic acid)
d- Laisser sécher pendant 5 minutes
e- Introduire la plaque dans l'instrument MALDI-TOF
f- Lancer l'analyse

Dans cette revue, il est question de donner les stratégies et de soulever toutes les contraintes liées à l'installation d'un instrument MALDI-TOF dans des pays en développement notamment en Afrique et

dans les pays tropicaux. Dans ces pays les contraintes financières et techniques sont les plus récurrentes pour l'acquisition d'un appareil comme le MALDI-TOF, mais toutefois il est judicieux de faire recours à des fondations et/ou à des organisations non gouvernementales afin de pouvoir supporter le coût d'investissement de l'appareil. En outre la mutualisation du MALDI-TOF MS a permis à plusieurs structures de santé telles que l'hôpital Le Dantec de Dakar, l'Institut de recherche pour le développement (IRD) de Dakar et l'Institut Pasteur (IP) de Dakar de bénéficier de l'utilisation de l'instrument installé à l'hôpital Principal de Dakar. Enfin, cette revue est une vision prospective pour l'avenir du MALDI-TOF MS dans les laboratoires de microbiologie clinique, qui consiste à précipiter le diagnostic et l'identification des échantillons cliniques et à mieux prendre en charge le patient.

MiniReview

Value of MALDI-TOF MS in Clinical Microbiology and Infectious Diseases in Africa and tropical areas

Cheikh Ibrahima Lo[1], Bécaye Fall[2], Bissoume Sambe-Ba[2], Christophe Flaudrops[1], Ngor Faye[3], Oleg Mediannikov[1], Cheikh Sokhna[1], Boubacar Wade[2], Didier Raoult[1], Florence Fenollar[1*]

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Marseille, France and Dakar, Senegal

[2]Hôpital Principal de Dakar, Senegal

[3]Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann, Senegal

*Corresponding author
E-mail: florence.fenollar@univ-amu.fr
Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
Tel: +33 491 385 517, Fax: +33 491 387 772

Abstract word count: 144
Text word count: 2.208

Keywords: MALDI-TOF MS, Africa, bacteria, clinical microbiology, infection diseases, hospital, Senegal.

Abstract

MALDI-TOF MS is a revolutionary technique with multiple applications. Currently, its use in clinical microbiology laboratory becomes widespread as the method is easy, rapid, effective, accuracy, and cheap for identification of cultured bacteria and fungi. Thus, it is an ideal tool to replace conventional methods still used in Africa and tropical areas for the microbiological routine diagnosis. Besides, the recent installation of a MALDI-TOF MS for diagnosis purpose in a hospital in Senegal has confirmed that this tool is valuable but also robust in tropical Africa supporting that this technique should be widely distributed there. However, despite its value for clinical microbiology laboratories in Africa, the acquisition and installation of MALDI-TOF MS is subject to several constraints. This review provides general aspects of MALDI-TOF MS. The specific aspects and constraints observed in Africa and tropical countries are also addressed with suggestions of adapted solutions.

Introduction

In developed countries, cardiovascular diseases are the leading cause of death, whereas in Africa and low-income countries, thousands of deaths linked to infectious diseases are recorded every year [1; 2; 3; 4]. In this context of a high incidence of infectious diseases, including emerging and reemerging pathogens [5], improvement of tools for the identification of microorganisms is required in clinical microbiology laboratories. In African countries, routine diagnostic methods are generally based on culture media followed by growth characteristics and biochemical patterns. These steps are fastidious, requiring a lot of expensive reagents and an *a priori* knowledge of the isolated microorganism; identification is done after many hours or days depending on the microorganisms and is sometimes inaccurate [6].

Recently, Matrix-Assisted Laser Desorption Ionization Time-Of-Flight (MALDI-TOF) mass spectrometry (MS) has enabled a revolution in the routine work of clinical microbiology laboratories with the quick, inexpensive, and accurate identification of bacteria and fungi without *a priori* knowledge [7]. Currently, the main cost of MS is that of the apparatus. MALDI-TOF MS has a real impact on global health and its implementation will be of great value in Africa and other tropical areas, as recently shown in Senegal, where its broad applicability and robustness have been recently proven [8].

Here, we will review the general aspects of MALDI-TOF MS but we will also focus on the specific aspects and constraints observed in Africa and tropical countries and we will propose adapted solutions.

General aspects of MALDI-TOF MS technology

In 1975, the scientific literature began to associate MS with pyrolysis for the detection of bacterial proteins [9]. In 2009, a new revolution in clinical microbiology was initiated when the efficiency of MALDI-TOF MS for the routine identification of bacteria was demonstrated with a correct identification of 95% and 84% of the genus and species levels, respectively, for 1,660 bacteria [6; 10]. Since then, an explosion of scientific publications on the use of MALDI-TOF MS in clinical microbiology has been observed (Figure 1) supporting the fact that the method is fast and reliable for the identification of microorganisms and clearly more efficient than conventional methods [11; 12]. It has been evaluated that ten bacterial strains can be identified in parallel in less than 15 minutes with MS whereas more than 360 minutes are required on conventional automated systems [13; 14].

Currently, three MALDI-TOF mass spectrometers are commercialized (Figure 2): Andromas system (Paris, France) [15], Microflex LT (Bruker Daltonics, Heidelberg, Germany, in collaboration with Becton Dickinson, Franklin Lakes, NJ, USA) [16; 17], and VITEK® MS (bioMérieux, Marcy l'Etoile, France) [18]. For bacterial or fungal identification, one isolated colony is picked and

directly deposited on a well of a MALDI-TOF plate, preferentially in duplicate as deposit is crucial for accurate identification (Figure 3) [19]. Then, this preparation must be overlaid with a matrix solution (solution with alpha-cyano-4-hydroxycinnamic acid, acetonitrile, trifluoroacetic acid, etc.), and air dried at room temperature (about 5 minutes) to permit co-crystallization [20] before introducing the plate in the MALDI-TOF instrument for analysis. The identification is achieved by comparing the spectra of analyzed species against the reference spectra present in the MALDI-TOF database [21; 22]. Identification robustness depends on the richness of the databases, which have been regularly and greatly updated since 2009 [10].

Recently, Tran *et al.* performed a huge study to evaluate cost savings realized by implementing routine microbiological identification by MALDI-TOF MS (bioMérieux Vitek, Durham, NC, USA) in their laboratory. Overall, reagent costs for the conventional methods averaged $3.59 per isolate and those for MS were $0.43. Thus, the use of MS equated to a net saving of $69,108.61 (87.8%) in reagent costs annually. When technologist time and maintenance costs were included, conventional identification would have cost $142,532.69 versus $68,886.51 with MS, resulting in a laboratory saving of $73,646.18 (51.7%) annually. They also estimated that the initial cost of the instrument at their usage level would be offset in about three years [23].

The direct identification of microorganisms in specimens such as blood cultures, urine, or cerebrospinal fluid has been proposed using

home-made [24; 25; 26;27; 28; 29] or commercial kits for blood cultures [30; 31; 32]. Currently, direct identification is mainly available for blood cultures after a pre-incubation step but allowing the quick identification of the involved microorganism and the opportunity of considerably earlier treatment adaptation with a direct clinical impact [33].

Although more complex than identification, as strain comparison requires statistical algorithms, hierarchical clustering, and a bioinformatics approach for the analyses of spectra and peaks, MALDI-TOF MS has been successfully used for microbial typing and identification at the subspecies level [34; 35; 36]. MALDI-TOF MS is a promising tool offering the potential to be used as an easy, inexpensive, and rapid epidemiological typing tool for outbreak investigation [34; 35; 36]. For example, the epidemiological investigation of a nosocomial outbreak of multidrug resistant *Corynebacterium striatum* has showed that all outbreak-related strains clustered in a single clone with a MALDI-TOF MS dendrogram [37]. This tool was also successfully utilized for differentiating *Streptococcus pneumoniae* outbreak isolates [38]. It has also allowed the accurate and reproducible discrimination of major methicillin-resistant *Staphylococcus aureus* (MRSA) clonal complexes observed in outbreaks, belonging to strains prepared with the same extraction protocol [39; 40]. *Listeria monocytogenes* serotypes have been correctly separated using discriminating peaks [41]. MALDI-TOF MS has also allowed the differentiation of the five most frequently isolated

157 *Salmonella enterica* serovars (Enteritidis, Typhimurium, Virchow,
158 Infantis, and Hadar) [42] as well as the determination of *Escherichia*
159 *coli* pathotypes [43]. Recently, the validity of MALDI-TOF MS for
160 typing extended-spectrum β-lactamase-producing *Escherichia coli* in a
161 previously published nosocomial outbreak was assessed [44].

162 MALDI-TOF MS also enabled the rapid detection of arthropod
163 vectors, such as ticks, mosquitoes, fleas [45], phlebotomine sand flies
164 [46], and *Culicoides* without expertise or skills in entomology [47;
165 48]. The utility of MALDI-TOF MS for a dual identification of tick
166 species and bacteria has been recently demonstrated. Indeed, detection
167 of intracellular *Rickettsia* spp. has been performed using MALDI-TOF
168 MS in ticks [49] as well as *Borrelia crocidurae* in *Ornithodoros*
169 *sonrai* ticks [50]. Therefore, MALDI-TOF MS seems to be a relevant
170 tool for the accurate and rapid identification of vectors and vectorized
171 pathogens for both field monitoring and entomological diagnosis. This
172 concept offers new perspectives for the monitoring of other vector
173 borne diseases that present public health concerns. Indeed, this
174 approach could be used to distinguish between uninfected mosquitoes
175 and those infected with sporozoites of *Plasmodium* spp. and to
176 determine potential vector resistance to insecticides in tropical areas.
177 Finally, MALDI-TOF MS also enabled the determination of meat
178 origin in raw and processed meat in culinary preparations, as well as
179 fish [51; 55]. Milestones of the use of MALDI-TOF MS for
180 identification purposes other than microbial are summarized in Table
181 1 [45; 48; 51; 52; 53; 54; 55].

Value of MALDI-TOF MS in Africa

Bacterial and fungal identification in Africa

Conventional biochemical identification methods (Figure 4) are usually performed in Africa but sometimes there are performance limitations [18; 56]. Indeed, storage of the various reagents requires strict conditions and compliance with the expiration date. However, problems of cold storage are observed with a real impact on reagents. Reagent supply issues are also observed. Finally, the identification is often based on the interpretations of the few biochemical tests available leading to an inaccurate identification with a clinical impact on patient management. Thus, use of new-generation technologies such as MALDI-TOF MS can resolve many of these difficulties.

The low cost, speed, and accuracy of identification without prior knowledge supports the fact that use of MALDI-TOF MS will help in microbiology laboratories in Africa [14; 57]. When a mass spectrometer was implemented in Senegal (Hôpital Principal de Dakar) in 2012, conventional methods such as API strips were immediately dropped. In just ten months, the instrument has correctly identified 2,082 bacteria and fungi at the level species (85.7%) [58].

Specific aspects and constraints for MALDI-TOF MS in Africa

Constraints for acquisition and installation

The primary obstacle to performing microbial identification using MALDI-TOF MS is the cost of the apparatus, which is estimated at between $120,000 and $270,000 [23].

Electricity is another constraint as it must be supplied continuously for MS. Thus, the presence of an electric generator is required to prevent power failure. Moreover, the instrument as well as all the connected computers must be equipped with an inverter in case of micro-power cuts. The room for the apparatus must be protected from insects and dust and thermo-isolated; air conditioning is mandatory.

Constraints for routine microbial identification

The main reagent required to perform MALDI-TOF MS is the chemical matrix, which is not expensive, in particular when home-made [6; 59]. Home-made solutions can be freshly prepared each day in no more than ten minutes and stored at room temperature for the day. When purchased or prepared a few days before use, the matrix must be stored at +4°C. Thus, reagents are not a limitation for the process of microbial identification. Each system includes spot target plates, but the plates are reusable steel targets for the Microflex LT while the VITEK® MS uses disposable plastic targets [60].

Humans may be a constraint as the personnel should be previously and specifically trained in the use of MALDI-TOF MS, but it is an easy system that does not require previous specific expertise. Besides, the required skills are quickly acquired. For example, in Senegal, after a 4-day course, including theoretical and practical training, all four people who completed the training course were autonomous in the use of MALDI-TOF MS.

Constraints for maintenance

The second main obstacle to the use of MALDI-TOF MS in Africa is maintenance. Indeed, maintenance once a year is recommended by the manufacturers. Maintenance raises two problems: its cost, including the cost of spare parts and labor or that of maintenance contracts; and the lack of trained personnel in Africa to perform it. The spare parts that need to be changed most frequently are the laser and detector (depending on frequency of use) and the primary pump (life of three to four years).

Solutions

Funding for the acquisition and maintenance of MALDI-TOF MS in Africa is the main constraint for implementing the technology, whereas routine identification does not actually raise problems or limitations. For us, the cost of acquiring the apparatus in Dakar was covered and shared between several organizations, including the Institute of Research for Development, a public French organization involved in research with and for southern countries, the Mediterranean Infection Hospital-University Institute, which promotes the fight against infectious diseases on a global scale, and the French Ministry of Foreign Affairs [8]. For others, research organizations, non-governmental organizations, or charity foundations, such as the Mérieux Foundation or the Melinda and Bill Gates Foundation, which are both already involved in the implementation of new tools to

prevent and treat deadly diseases in Africa, could help fund the mass spectrometer.

Currently, the strategy applied in several countries in order to lower management costs is the grouping of clinical microbiology laboratories into large core laboratories. Thus, the development of a common MS platform between several clinical microbiology laboratories in nearby areas seems the best option in order to share the costs. In Dakar, in Senegal, the MALDI-TOF MS located in the Hôpital Principal is also open to other structures for research and diagnostic purposes (Figure 5). The experience of MALDI-TOF MS networking in university hospitals in Belgium has been recently reported for identification of microorganisms in Brussels [59]. Indeed, during a 1-month period, 1,055 isolates were identified by conventional techniques in a first hospital and analyzed by MALDI-TOF MS in another hospital situated 7.5 km away; target plates and identification projects were sent. The median time to identification was 5h 11 min [61]. Identification by the MALDI-TOF networking system was more accurate and faster than that carried out in parallel with conventional methods and led to substantial annual cost savings [59]. Our clinical microbiology laboratory (University Hospital, Marseille, France) also opened up access 12 months ago to our MALDI-TOF MS platform for use by other hospitals: the public health hospital from Salon de Provence, a remote town 52 kilometers away with 400 beds, and the military teaching hospital of Marseille (Laveran), a general hospital with open access for both military and

civilians with 303 beds (personal data). Every week, hundreds of bacteria were correctly identified at a low cost without moving patients. Thus, use of a same MALDI-TOF MS platform allows the sharing of competence and a decrease in the cost of acquiring and maintaining the instrument [59].

If a MALDI-TOF MS platform is established, a cooperation agreement and a convention should be established between the different teams in order to specify not only the organization of workflows but also the tasks and responsibilities of each one. Finally, for maintenance, local staff should be specifically trained.

Conclusion and perspectives

The rapidity, efficiency, and low cost have led many laboratories to adopt the MALDI-TOF as a tool for routine diagnosis, resulting in an improvement in patient care. The first successful implementation of a MALDI-TOF mass spectrometer in Senegal supports the fact that it is a robust and valuable tool in tropical Africa and should be widely distributed there. The development of common MALDI-TOF MS platforms in nearby geographic areas will allow the sharing of equipment, skills, and costs.

Acknowledgements.

We thank the Mediterranean Infection Hospital-University.

Legend figures.

Figure 1. Increasing number of publications related to MALDI-TOF MS applications in medical microbiology from 1999 to 2014. It shows also that 2009 marks the massive use of MALDI-TOF MS in clinical microbiology laboratories. The following Mesh term through bibliographic database NCBI were used to built this graph: ("spectrometry, mass, matrix-assisted laser desorption-ionization"[MeSH Terms] OR ("spectrometry"[All Fields] AND "mass"[All Fields] AND "matrix-assisted"[All Fields] AND "laser"[All Fields] AND "desorption-ionization"[All Fields]) OR "matrix-assisted laser desorption-ionization mass spectrometry"[All Fields] OR "maldi"[All Fields]) AND tof[All Fields]) AND ("microbiology"[Subheading] OR "microbiology"[All Fields] OR "microbiology"[MeSH Terms]) AND ("1999/01/01"[PDAT] : "2014/12/31"[PDAT]).

Figure 2. The different MALDI-TOF MS instruments currently commercialized for identification of microorganisms in clinical laboratory.

LT2-Andromas, Vitek MS, and MALDI Biotyper were accredited for identification purposes in clinical microbiology laboratory according to EU directive EC/98/79 in several European countries. VITEK® MS and tMALDI Biotyper have been cleared by the US Food and Drug Administration (FDA) for identification of cultured bacteria and, in the case of the former system, yeast in 2013.

Figure 3. MALDI-TOF MS's operating principle and the sample preparation step for identification. The principle of this measurement is based on the ability of an electric and / or magnetic field deflecting a flow of ions, each with a mass and a charge proportional for their trajectories. The mass spectrometry can be overall divided in three steps: the ionization chamber that produces ions in the gas phase (A), the analyzer which selects ions by mass to charge ratio (m/z) (B), and the detector that converts the ionic current in electric current (C). A bombing of a laser beam allows generating ions in the ionization chamber. These ions were accelerated in an electric field which directs them to the analyzer which separates them according to their time of free flight (TOF: Time-Of-Flight). The smaller molecules reached in first the detector following the biggest, according to the m/z ratio. Then those which have the same ratio m/z are separated by an electrostatic mirror. The detector converts the received ions in electrical current which is amplified and digitized (D).

Figure 4. MALDI-TOF MS performance compared to conventional methods routinely used in some clinical laboratories in Africa.

Figure 5. Schematization of the circuit for the use of a MALDI-TOF platform in Dakar (Senegal). MALDI-TOF MS is located in the clinical microbiology laboratory of the Hôpital Principal de Dakar (A), but it is also used for diagnostic and research purposes by the Hospital Le Dantec and the Institute Pasteur of Dakar (B) and by the Institut de Recherche pour le Développement (IRD) situated at Eastern of Dakar (C).

References

1. Prost A. « Maladies infectieuses : nouveau destin, nouveaux concepts ». Espace, populations, sociétés 18, no 2 (2000): 159‑65. doi:10.3406/espos.2000.1938.
2. Lopez AD, Ahmad OB, Guillot M, Ferguson BD, Salomon JA, Murray CJL and, Hill KH. World Mortality in 2000: Life Tables for 191 Countries. Geneva: WHO 2002.
3. Bryce J, Boschi-Pinto C, Shibuya K, Black RE, WHO Child Health Epidemiology Reference Group. WHO estimates of the causes of death in children. *Lancet*. 2005; 365(9465):1147‑52.
4. Williams BG, Gouws E, Boschi-Pinto C, Bryce J, Dye C. Estimates of world-wide distribution of child deaths from acute respiratory infections. *Lancet Infect Dis*. 2002; 2(1):25‑32.
5. Desenclos JC, De Valk H. [Emergent infectious diseases: importance for public health, epidemiology, promoting factors, and prevention]. *Med Mal Infect*. 2005; 35(2):49‑61.
6. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis*. 2009; 49(4):543‑51.
7. Courcol R. Quelles utilisations de la spectrométrie de masse de type MALDI-TOF en microbiologie médicale ? *Revue Francophone des Laboratoires*. 2009; (416):61‑4.
8. Fall B, Lo CI, Samb-Ba B, Perrot N, Diawara S, Gueye MW, et al. The ongoing revolution of maldi-tof mass spectrometry for microbiology reaches tropical Africa. *Am J Trop Med Hyg*. 2015; 92(3):641‑7.
9. Intelicato-Young J, Fox A. Mass spectrometry and tandem mass spectrometry characterization of protein patterns, protein markers and whole proteomes for pathogenic bacteria. *J Microbiol Methods*. 2013;92(3):381‑6.
10. Seng P, Rolain JM, Fournier PE, La Scola B, Drancourt M, Raoult D. MALDI-TOF-mass spectrometry applications in clinical microbiology. *Future Microbiol*. 2010; 5(11):1733‑54.

11. Eigner U, Holfelder M, Oberdorfer K, Betz-Wild U, Bertsch D, Fahr AM. Performance of a matrix-assisted laser desorption ionization-time-of-flight mass spectrometry system for the identification of bacterial isolates in the clinical routine laboratory. *Clin Lab.* 2009; 55(7-8):289‑96.

12. Blondiaux N, Gaillot O, Courcol R-J. [MALDI-TOF mass spectrometry to identify clinical bacterial isolates: evaluation in a teaching hospital in Lille]. *Pathol Biol*. 2010; 58(1):55-7.

13. Biswas S, Rolain JM. Use of MALDI-TOF mass spectrometry for identification of bacteria that are difficult to culture. *J Microbiol Methods*. 2013; 92(1):14‑24.

14. Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, Francois P, et al. Comparison of two matrix-assisted laser desorption ionization time-of-flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *J Clin Microbiol*. 2010; 48(4):1169‑75.

15. Bille E, Dauphin B, Leto J, Bougnoux M-E, Beretti J-L, Lotz A, et al. MALDI-TOF MS Andromas strategy for the routine identification of bacteria, mycobacteria, yeasts, *Aspergillus* spp. and positive blood cultures. *Clin Microbiol Infect.* 2012; 18(11):1117‑25.

16. Lee M, Chung H-S, Moon H-W, Lee SH, Lee K. Comparative evaluation of two matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) systems, Vitek MS and Microflex LT, for the identification of Gram-positive cocci routinely isolated in clinical microbiology laboratories. *J Microbiol Methods*. 2015; 113:13‑5.

17. Saffert RT, Cunningham SA, Ihde SM, Jobe KE, Mandrekar J, Patel R. Comparison of Bruker Biotyper matrix-assisted laser desorption ionization time-of-flight mass spectrometer to BD Phoenix automated microbiology system for identification of gram-negative bacilli. *J Clin Microbiol*. 2011; 49(3):887-92.

18. Patel R. Matrix-assisted laser desorption ionization-time of flight mass spectrometry in clinical microbiology. *Clin Infect Dis.* 2013; 57(4):564-72.

19. Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom Rev*. 2001; 20(4):157‑71.

20. Shunsuke S, Setsu S, Shota M, Motoi N, Minako, Kazuyuki S, Masaharu W, Mamoru S, Tomoo M, Masayoshi K, Yasuo I, Satoshi K, Naokatsu S, Fumio N. Direct application of MALDI-TOF mass spectrometry to cerebrospinal fluid for rapid pathogen identification in a patient with bacterial meningitis. *Clinica Chimica Acta.* 2014; 435:59–61.

21. Coltella L, Mancinelli L, Onori M, Lucignano B, Menichella D, Sorge R, et al. Advancement in the routine identification of anaerobic bacteria by MALDI-TOF mass spectrometry. *Eur J Clin Microbiol Infect Dis*. 2013; 32(9):1183‑92.

22. Martiny D, Busson L, Wybo I, El Haj RA, Dediste A, Vandenberg O. Comparison of the Microflex LT and Vitek MS systems for routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *J Clin Microbiol*. 2012; 50(4):1313-25.

23. Tran A, Alby K, Kerr A, Jones M, Gilligan PH. Cost Savings Realized by Implementation of Routine Microbiological Identification by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. *J Clin Microbiol*. 2015;53(8):2473−9.

24. Segawa S, Sawai S, Murata S, Nishimura M, Beppu M, Sogawa K, Watanabe M, Satoh M, Matsutani T, Kobayashi M, Iwadate Y, Kuwabara S, Saeki N, Nomura F. Direct application of MALDI-TOF mass spectrometry to cerebrospinal fluid for rapid pathogen identification in a patient with bacterial meningitis. *Clin Chim Acta*. 2014;435:59-61.

25. Ferreira L, Sánchez-Juanes F, González-Avila M, Cembrero-Fuciños D, Herrero-Hernández A, González-Buitrago JM, Muñoz-Bellido JL. Direct identification of urinary tract pathogens from urine samples by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*. 2010; 48(6):2110-5.

26. Ferreira L, Sánchez-Juanes F, Muñoz-Bellido JL, González-Buitrago JM. Rapid method for direct identification of bacteria

in urine and blood culture samples by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: intact cell vs. extraction method. *Clin Microbiol Infect*. 2011; 17(7):1007‑12.

27. La Scola B, Raoult D. Direct identification of bacteria in positive blood culture bottles by matrix-assisted laser desorption ionisation time-of-flight mass spectrometry. *PLoS One*. 2009; 4(11):e8041.

28. Foster AW. Rapid identification of microbes in positive blood cultures by matrix-assisted laser desorption/ionisation Time of flight (MALDI-TOF) mass spectrometry (Vitek-MS-bioMérieux). *J Clin Microbiol.* 2013; 51 (11): 3717-3719.

29. Ferroni A, Suarez S, Beretti JL, Dauphin B, Bille E, Meyer J, et al. Real-time identification of bacteria and Candida species in positive blood culture broths by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *J Clin Microbiol*. 2010; 48(5):1542‑8.

30. Jamal W, Saleem R, Rotimi VO. Rapid identification of pathogens directly from blood culture bottles by Bruker matrix-assisted laser desorption laser ionization time-of-flight mass spectrometry versus routine methods. *Diagn Microbiol Infect Dis*. 2013; 76(4):404-8.

31. Haigh JD, Green IM, Ball D, Eydmann M, Millar M, Wilks M. Rapid identification of bacteria from bioMérieux BacT/ALERT blood culture bottles by MALDI-TOF MS. *Br J Biomed Sci.* 2013;70(4):149-55.

32. Nonnemann B, Tvede M, Bjarnsholt T. Identification of pathogenic microorganisms directly from positive blood vials by matrix-assisted laser desorption/ionization time of flight mass spectrometry. *APMIS*. 2013; 121(9):871‑7.

33. Kohlmann R, Hoffmann A, Geis G, Gatermann S. MALDI-TOF mass spectrometry following short incubation on a solid medium is a valuable tool for rapid pathogen identification from positive blood cultures. *Int J Med Microbiol*. 2015.

34. Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiol Rev*. 2012, 36 (2):380-407.

35. Gaia V, Casati S, Tonolla M. Rapid identification of *Legionella spp*. by MALDI-TOF MS based protein mass fingerprinting. *Syst Appl Microbiol*. 2011;34(1):40‑4.

36. Fujinami Y, Kikkawa HS, Kurosaki Y, Sakurada K, Yoshino M, Yasuda J. Rapid discrimination of *Legionella* by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Microbiol Res*. 2011; 166(2):77-86.

37. Verroken A, Bauraing C, Deplano A, Bogaerts P, Huang D, Wauters G, et al. Epidemiological investigation of a nosocomial outbreak of multidrug-resistant *Corynebacterium striatum* at one Belgian university hospital. *Clin Microbiol Infect*. 2013; 20(1):44–50.

38. Williamson YM, Moura H, Woolfitt AR, Pirkle JL, Barr JR, Carvalho Mda G, et al. Differentiation of *Streptococcus pneumoniae* conjunctivitis outbreak isolates by matrix-assisted laser desorption ionization time of flight mass spectrometry. *Appl Environ Microbiol.* 2008; 74(19):5891-7.

39. Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, et al. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin resistant Staphylococcus aureus lineages. *Int J Med Microbiol.* 2011; 301(1):64–8.

40. Josten M, Reif M, Szekat C, Al-Sabti N, Roemer T, Sparbier K, et al. Analysis of the matrix-assisted laser desorption ionization time of flight mass spectrum of Staphylococcus aureus identifies mutations that allow differentiation of the main clonal lineages. *J Clin Microbiol.* 2013; 51(6):1809–17.

41. Barbuddhe SB, Maier T, Schwarz G, Kostrzewa M, Hof H, Domann E, et al. Rapid identification and typing of *listeria* species by matrix-assisted laser desorption ionization time of flight mass spectrometry. *Appl Environ Microbiol.* 2008; 74(17):5402–7.

42. Dieckmann R, Malorny B. Rapid screening of epidemiologically important *Salmonella enterica* subsp. *enterica* serovars by whole cell matrix-assisted laser desorption ionization time of flight mass spectrometry. *Appl Environ Microbiol.* 2011; 77(12):4136–46.

538 **43.** Clark CG, Kruczkiewicz P, Guan C, McCorrister SJ, Chong P,
539 Wylie J, et al. Evaluation of MALDI-TOF mass spectroscopy
540 methods for determination of *Escherichia coli* pathotypes. *J*
541 *Microbiol Methods*. 2013; 94(3):180‑91.
542 **44.** Egli A, Tschudin-Sutter S, Oberle M, Goldenberger D, Frei R,
543 Widmer AF. Matrix-assisted laser desorption/ionization time of
544 flight mass-spectrometry (MALDI-TOF MS) based typing of
545 extended-spectrum β-lactamase producing *E. coli* a novel tool
546 for real-time outbreak investigation. *PLoS One*.
547 2015;10(4):e0120624.
548 **45.** Yssouf A, Socolovschi C, Leulmi H, Kernif T, Bitam I, Audoly
549 G, et al. Identification of flea species using MALDI-TOF/MS.
550 *Comp Immunol Microbiol Infect Dis*. 2014; 37(3):153 7.
551 **46.** Mathis A, Depaquit J, Dvořák V, Tuten H, Bañuls A-L, Halada
552 P, et al. Identification of phlebotomine sand flies using one
553 MALDI-TOF MS reference database and two mass
554 spectrometer systems. *Parasit Vectors.* 2015; 8:266.
555 **47.** Sambou M, Aubadie-Ladrix M, Fenollar F, Fall B, Bassene H,
556 Almeras L, et al. Comparison of matrix-assisted laser
557 desorption ionization-time of flight mass spectrometry and
558 molecular biology techniques for identification of *Culicoides*
559 (*Diptera*: *ceratopogonidae*) biting midges in senegal. *J Clin*
560 *Microbiol*. 2015; 53(2):410‑8.
561 **48.** Yssouf A, Socolovschi C, Flaudrops C, Ndiath MO,
562 Sougoufara S, Dehecq JS, et al. Matrix-assisted laser desorption
563 ionization time-of-flight mass spectrometry: an emerging tool
564 for the rapid identification of mosquito vectors. *PLoS One*.
565 2013;8(8):e72380.
566 **49.** Yssouf A, Almeras L, Terras J, Socolovschi C, Raoult D,
567 Parola P. Detection of *Rickettsia* spp in ticks by MALDI-TOF
568 MS. *PLoS Negl Trop Dis*. 2015;9(2):e0003473.
569 **50.** Fotso Fotso A, Mediannikov O, Diatta G, Almeras L, Flaudrops
570 C, Parola P, et al. MALDI-TOF mass spectrometry detection of
571 pathogens in vectors: the *Borrelia crocidurae*/Ornithodoros
572 sonrai paradigm. *PLoS Negl Trop Dis.* 2014;8(7):e2984.

573 **51.** Mazzeo MF, Giulio BD, Guerriero G, Ciarcia G, Malorni A,
574 Russo GL, et al. Fish authentication by MALDI-TOF mass
575 spectrometry. *J Agric Food Chem*. 2008;56(23):11071 6.
576 **52.** Kaufmann C, Schaffner F, Ziegler D, Pflüger V, Mathis A.
577 Identification of field-caught Culicoides biting midges using
578 matrix-assisted laser desorption/ionization time of flight mass
579 spectrometry. *Parasitology*. 2012;139(2):248 58.
580 **53.** Yssouf A, Flaudrops C, Drali R, Kernif T, Socolovschi C,
581 Berenger J-M, et al. Matrix-assisted laser desorption ionization-
582 time of flight mass spectrometry for rapid identification of tick
583 vectors. *J Clin Microbiol*. 2013;51(2):522 8.
584 **54.** Steinmann IC, Pflüger V, Schaffner F, Mathis A, Kaufmann C.
585 Evaluation of matrix-assisted laser desorption/ionization time
586 of flight mass spectrometry for the identification of
587 ceratopogonid and culicid larvae. *Parasitology*.
588 2013;140(3):318‑27.
589 **55.** Flaudrops C, Armstrong N, Raoult D, Chabrière E.
590 Determination of the animal origin of meat and gelatin by
591 MALDI-TOF-MS. *J Food Compost Anal*. 2015; 41:104‑12.
592 **56.** Samb-Ba B, Mazenot C, Gassama-Sow A, Dubourg G, Richet
593 H, Hugon P, Lagier JC, Raoult D, Fenollar F. MALDI-TOF
594 identification of the human Gut microbiome in people with and
595 without diarrhea in Senegal. *PLoS One*. 2014; 9(5):e87419.
596 **57.** Bizzini A, Greub G. Matrix-assisted laser desorption ionization
597 time-of-flight mass spectrometry, a revolution in clinical
598 microbial identification. *Clin Microbiol Infect*. 2010;
599 16(11):1614‑9.
600 **58.** Moussaoui L. Applications de la spectrométrie de masse type
601 MALDI-TOF à la bactériologie et à la distinction de variants
602 génétiques [Internet] [phdthesis]. Université de Strasbourg;
603 2012 [cited 15 Apr. 2015]. Available on: https://tel.archives-
604 ouvertes.fr/tel-00872251/document
605 **59.** Martiny D, Cremagnani P, Gaillard A, Miendje Deyi VY,
606 Mascart G, Ebraert A, Attalibi S, Dediste A, Vandenberg O.
607 Feasibility of matrix-assisted laser desorption/ionisation time-
608 of-flight mass spectrometry (MALDI-TOF MS) networking in

university hospitals in Brussels. *Eur J Clin Microbiol Infect.* 2014; 33:745-54.

60. Deak E, Charlton CL, Bobenchik AM, Miller SA, Pollett S, McHardy IH, et al. Comparison of the Vitek MS and Bruker Microflex LT MALDI-TOF MS platforms for routine identification of commonly isolated bacteria and yeast in the clinical microbiology laboratory. *Diagn Microbiol Infect Dis.* 2015; 81(1):27‑33.

61. Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clin Microbiol Rev*. 2015; 28(1):237‑64.

Table 1. Milestones of the use of MALDI-TOF MS for identification purposes other than microbial.

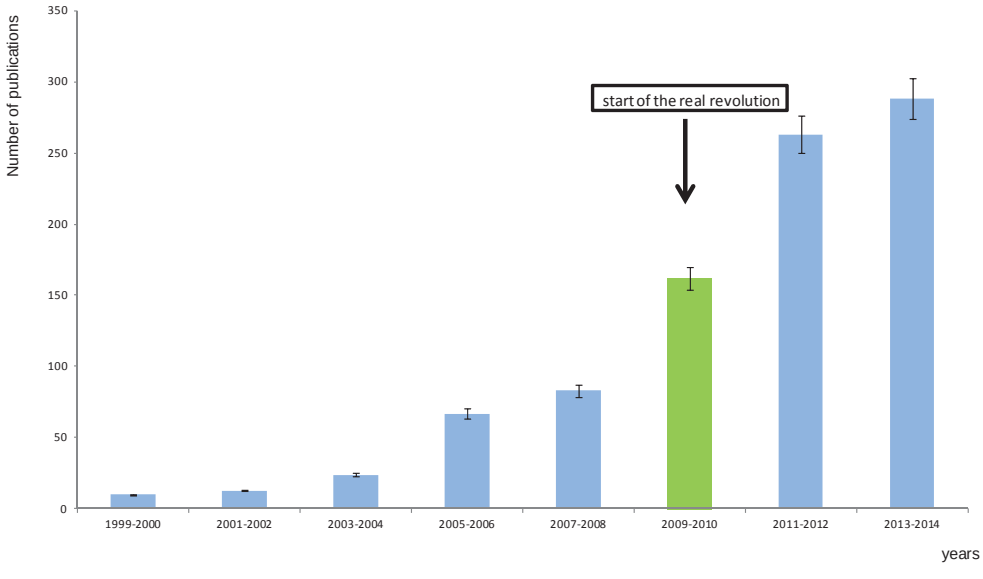
Author	First use of MALDI-TOF MS for identification purposes	Year
Mazzaeo et al [51]	Fish	2008
Kaufmann et al [52]	Culicoides	2012
Yssouf et al [53]	Ticks	2013
Steinmann et al [54]	Ceratopogonid and culicid larvae	2013
Yssouf et al [48]	Mosquitoes	2013
Yssouf et al [45]	Fleas	2014
Flaudrops et al [55]	Meat from raw and processed meat in culinary preparations	2015

628 **Figure 1.** Increasing number of publications related to MALDI-TOF
629 MS applications in medical microbiology from 1999 to 2014.

630

631

632



633

Figure 2. The different MALDI-TOF MS instruments currently commercialized for identification of microorganisms in clinical laboratory

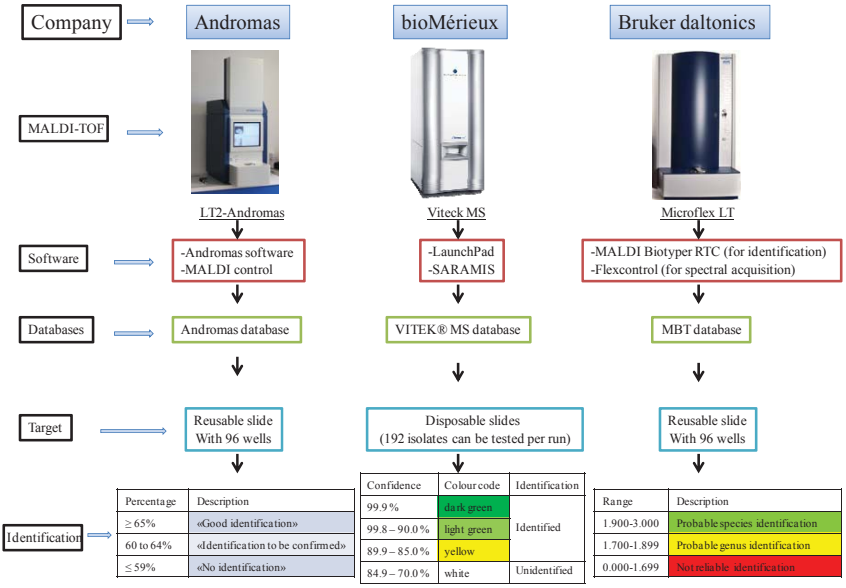


Figure 3. MALDI-TOF MS's operating principle and the sample preparation step for identification

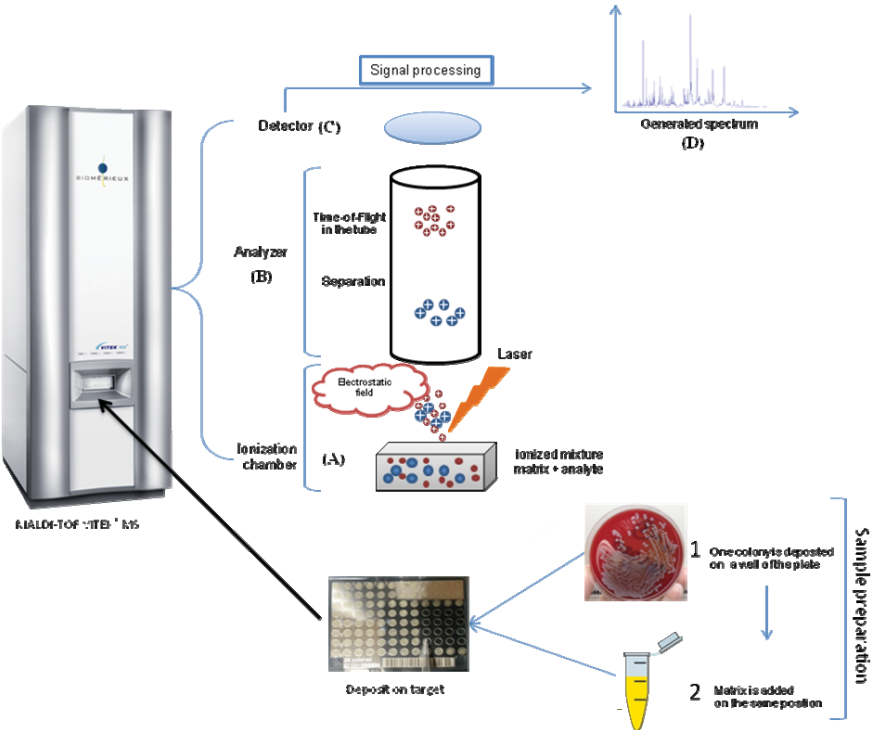


Figure 4. MALDI-TOF MS performance compared to conventional methods routinely used in some clinical laboratories in Africa.

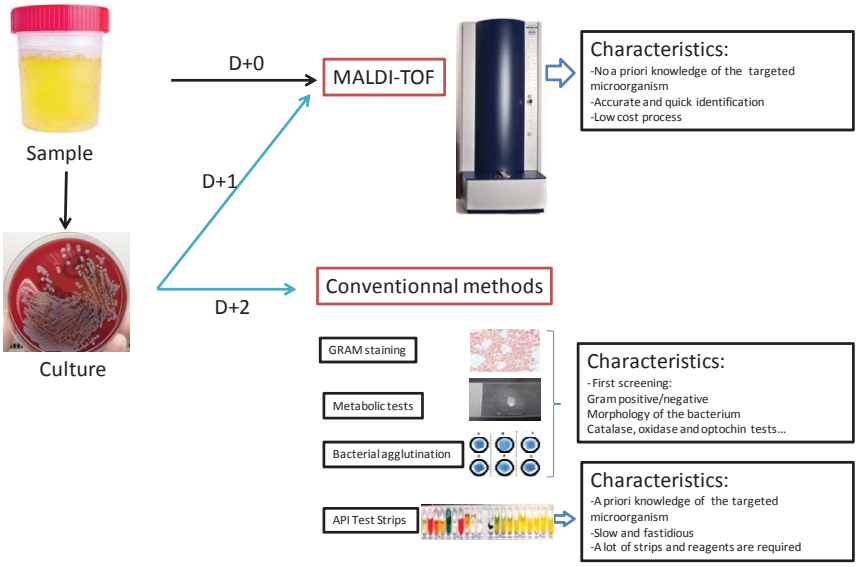
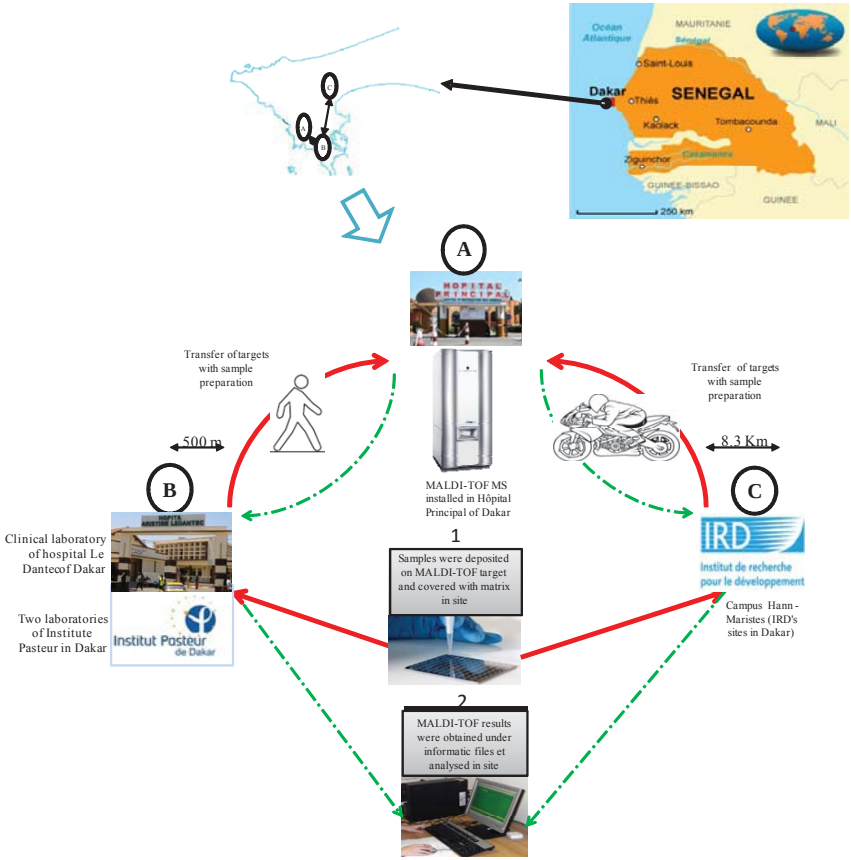


Figure 5. Schematization of the circuit for the use of a MALDI-TOF platform in Dakar (Senegal).



Conclusion et Perspective tirées du Chapitre 1

La spectrométrie de masse de type MALDI-TOF est une technique révolutionnaire récemment introduite dans les laboratoires de microbiologie clinique. Dans certains laboratoires, cette technologie est érigée comme outil de première ligne pour l'identification de routine des bactéries et levures isolées des prélèvements. L'acquisition d'un MALDI-TOF MS malgré son prix exorbitant a été possible en Afrique de l'ouest particulièrement au Sénégal, grâce à une importante aide financière émanant de structures étrangères.

En réalité dans les pays où il y a une forte prévalence des maladies infectieuses comme en Afrique et en zone tropicale, des outils permettant l'identification rapide et à haut débit des microorganismes, deviennent presque une nécessité pour une meilleure et précoce prise en charge des patients.

Dans l'avenir nous préconisons l'installation de MALDI-TOF MS dans plusieurs centres de santé à travers les grandes villes africaines. Par ailleurs une nouvelle stratégie qui consiste à mettre un MALDI-TOF MS à la disposition de plusieurs structures sanitaires situées dans une même zone géographique permet de faire des économies pour la maintenance, l'achat des consommables et du nombre de personnes recrutées pour ce travail. Le MALDI-TOF pourra être dans l'avenir un outil efficace pour explorer d'autres domaines de la microbiologie en Afrique, autrement l'identification de microorganismes d'intérêt médical comme les mycobactéries, les shigelles, et les champignons. En outre l'implantation en masse du MALDI-TOF permettra, l'identification d'autres pathogènes tels que les virus qui posent des problèmes de santé publique et de diagnostic dans les pays tropicaux

Chapitre 2:

Performance et faisabilité du MALDI-TOF MS en Afrique

Article 2:

« MALDI-TOF mass spectrometry: a powerful tool for clinical microbiology laboratory in Africa »

Cheikh Ibrahima Lo[1], Bécaye Fall[2], Bissoume Samb-Ba[2], Nadine Perrot[3], Silman Diawara[2], Mamadou Wague Gueye[2], Kowry Sow[2], Oleg Mediannikov[1], Cheikh Sokhna[1], Yaya Diemé[2], Sonia Chatellier[3], Boubacar Wade[2], Didier Raoult[1], Florence Fenollar[1].

[1]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CRNS7278, IRD198, InsermU1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Aix-Marseille Université, Marseille, France and Dakar, Senegal

[2]Hôpital Principal de Dakar, Senegal

[3]bioMérieux, La Balme-les Grottes, France

Soumis à *PloS ONE*, en cours de révision

Résumé de l'article 2: ***« MALDI-TOF mass spectrometry: a powerful tool for clinical microbiology laboratory in Africa»***

Contexte:

L'équipe de l'Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes à Marseille dans laquelle nous avons effectué nos travaux de recherche a été un pionnier dans la révolution pour l'identification de routine des microorganismes isolés dans les laboratoires cliniques par l'utilisation de la spectrométrie de masse MALDI-TOF (30–33). En 2009, elle a rapporté le premier travail d'identification des bactéries par MALDI-TOF au sein de son laboratoire de Microbiologie Clinique dans les Hôpitaux de Marseille en France. Cette équipe a ensuite participé en 2012 à l'installation d'un spectromètre de masse (SM) MALDI-TOF à l'hôpital Principal de Dakar au Sénégal afin d'évaluer la performance d'un tel appareil en Afrique tropicale.

Matériel et Méthode:

L'appareil installé à Marseille est un MALDI-TOF Microflex LT (Bruker Daltonics). C'est avec cet appareil que l'identification de routine des bactéries et levures isolées à partir des prélèvements des patients des hôpitaux de Marseille est réalisée. L'appareil installé à Dakar est un MALDI-TOF Vitek® MS (bioMérieux, Marcy l'Etoile, France). Cet appareil a été utilisé pour l'identification des bactéries présentes dans les prélèvements des patients de l'hôpital Principal de Dakar.

Dans ce travail, nous avons évalué et comparé l'identification de 190 souches isolées à partir des prélèvements cliniques issus de patients hospitalisés ou consultant à l'hôpital Principal de Dakar (HPD). Toutes ces souches étaient inoculées sur différents milieux de cultures fabriqués localement (gélose au sang de cheval, gélose Mueller-Hinton, gélose trypticase soja, gélose MacConkey etc.). Ensuite les colonies obtenues sont passées au MALDI-TOF Vitek MS (bioMérieux) pour l'identification. Puis les microorganismes mal identifiés sont envoyés à Marseille (France) pour une vérification d'abord avec le MALDI-TOF Microflex LT (Bruker Daltonics). Si l'absence d'identification persiste, nous procéderons à un séquençage du gène ARN ribosomal 16S pour résoudre définitivement la discordance ou le manque d'identification.

Résultats:

Dans l'ensemble 190 souches ont été testées successivement avec les spectromètres de masse MALDI-TOF de Dakar et de Marseille. Ainsi 153 bactéries (80,5%) ont été identifiées correctement jusqu'au niveau de l'espèce et 174 (91,6%) seulement jusqu'au niveau du genre. Les 10 bactéries les plus couramment identifiées dans le laboratoire à Dakar représentent 94,2% de la totalité des souches. Celles-ci étaient identifiées correctement par le SM MALDI-TOF de Dakar comme étant *Escherichia coli, Klebsiella pneumoniae, Streptococcus agalactiae, Acinetobacter baumanii, Pseudomonas aeruginosa, Staphylococcus aureus, Staphylococcus haemolyticus, Enterobacter cloacae, Enterococcus faecalis* et *Staphylococcus epidermidis*. L'espèce *Achromobacter xylosoxidans* a été systématiquement confondue avec

Achromobacter denitrificans par le MALDI-TOF Vitek MS. D'autres espèces *Exiguobacterium aurantiacum* (2 souches) et *Kytococcus schroeterii* (1 souche) n'ont jamais pu être identifiées à Dakar. Par ailleurs, quelques soucis d'identification ont été notés pour certaines espèces de *Bacillus* et de Streptocoques aussi bien avec le MALDI-TOF à Dakar qu'avec celui de Marseille. Enfin le séquençage du gène de l'ARN ribosomal 16S nous a conduits à l'identification d'une nouvelle bactérie nommée *Necropsobacter massiliensis*.

Conclusion:

L'identification robuste et concordante obtenue à la fois par les MALDI-TOF de Dakar et de Marseille prouve que cette technique peut être utilisée comme outil de première ligne dans les laboratoires de microbiologie clinique partout dans le monde notamment en Afrique.

MALDI-TOF mass spectrometry: a powerful tool for clinical microbiology at Hôpital Principal de Dakar, Senegal (West Africa)

Cheikh Ibrahima Lo[1], Bécaye Fall[2], Bissoume Sambe-Ba[2], Silman Diawara[2], Mamadou Wague Gueye[2], Oleg Mediannikov[1], Cheikh Sokhna[1], Ngor Faye[3], Yaya Diemé[2], Boubacar Wade[2], Didier Raoult[1], Florence Fenollar[1]

[1]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM63, CNRS7278, IRD198, InsermU1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Aix-Marseille Université, Marseille, France and Dakar, Senegal

[2]Hôpital Principal de Dakar, Senegal

[3]Université Cheikh Anta Diop, Dakar, Senegal

Correspondence to Florence Fenollar (florence.fenollar@univ-amu.fr)

Running title: Mass spectrometry: a powerful tool for clinical microbiology in Africa

Abstract count word: 243

Text count word: 2,957

Keywords: Mass spectrometry, MALDI TOF MS, Bacterial identification, fungal identification

ABSTRACT

Our team in Europe has developed the routine clinical laboratory identification of microorganisms by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS). To evaluate the utility of MALDI-TOF MS in tropical Africa in collaboration with local teams, we installed an apparatus in the Hôpital Principal de Dakar (Senegal), performed routine identification of isolates, and confirmed or completed their identification in France. In the case of discordance or a lack of identification, molecular biology was performed. Overall, 153/191 (80.1%) and 174/191 (91.%) isolates yielded an accurate and concordant identification for the species and genus, respectively, with the 2 different MALDI-TOF MSs in Dakar and Marseille. The 10 most common bacteria, representing 94.2% of all bacteria routinely identified in the laboratory in Dakar (*Escherichia coli*, *Klebsiella pneumoniae*, *Streptococcus agalactiae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Staphylococcus haemolyticus*, *Enterobacter cloacae*, *Enterococcus faecalis*, and *Staphylococcus epidermidis*) were accurately identified with the MALDI-TOF MS in Dakar. The most frequent misidentification in Dakar was at the species level for *Achromobacter xylosoxidans*, which was inaccurately identified as *Achromobacter denitrificans*, and the bacteria absent from the database, such as *Exiguobacterium aurientacum* or *Kytococcus schroeteri*, could not be identified. A few difficulties were observed with MALDI-TOF MS for *Bacillus* sp. or oral streptococci. 16S rRNA sequencing identified a novel bacterium, "*Necropsobacter massiliensis*." The robust identification of microorganisms by MALDI-TOF MS in Dakar and Marseille demonstrates that MALDI-TOF MS can be used as a first-line tool in clinical microbiology laboratories in tropical countries.

INTRODUCTION

The routine identification of bacteria and fungi by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) has become a true revolution in clinical microbiology laboratories [1,2]. Indeed, this technique is a powerful and robust tool for accurately identifying bacteria and fungi in less than 1 hour without a priori knowledge of the type of microorganisms; it is also more cost-effective than current phenotypic testing methods, despite the initial cost of the instrument and the maintenance costs, and perhaps most importantly, it is easy to use [1,2]. The rapid and accurate identification of microorganisms is necessary for the management of infectious diseases, particularly for choosing effective antibiotic therapies, reducing costs, and shortening hospital stays. MALDI-TOF MS has already become a first-line tool for routine microbial identification in many laboratories in Europe, and its use has become widespread in other areas, including Asia and America [2–7].

Since 2009, our laboratory in Marseille, France, has been a pioneer in the identification of bacteria using MALDI-TOF MS in the clinical microbiology laboratory [5]. Thus, for 6 years, the identification of bacteria has been routinely performed with our MALDI-TOF MS platform, and our database has undergone continuous revision [5,6]. Based on our extensive experience, we aimed to perform routine identification in a clinical microbiology laboratory in tropical Africa. Bacterial identification in Africa is challenging because conventional phenotypic methods require many reagents with specific storage conditions and shelf lives, in addition to higher consumable costs. Thus optimal conditions for bacterial identification are often lacking, and misidentification is prevalent. In addition, in the case of misidentification, molecular methods that are complementary for microbial identification cannot be performed due not only to complicated procedures but also the need for specific materials, reagents, and automates

with very high cost. We believe that MALDI-TOF MS can be a great alternative for bacterial and fungal identification in Africa, when electricity is continuously available as well as air conditioning. Thus, we installed a MALDI-TOF MS platform in the Hôpital Principal de Dakar in Senegal [8].

Herein, we evaluated the potential for routine bacterial and fungal identification using MALDI-TOF MS in a clinical microbiology laboratory in Africa.

MATERIALS AND METHODS

Isolates

Two hundred clinical isolates obtained by conventional culture procedures in the clinical microbiology laboratory of the Hôpital Principal de Dakar (Senegal) and tested for identification using MALDI-TOF MS (VITEK® MS-RUO) were sent to the clinical microbiology laboratory of the University Hospital in Marseille to confirm or complete the identification performed in Africa using another mass spectrometer (MicroFlex LT, Bruker Daltonik). As each MALDI-TOF system has been regulatory approved, is commercialized, and currently used for the routine bacterial and fungal identification in clinical microbiology laboratories in several European countries and in USA, we speculate that when the identifications were concordant, they were accurate.

When isolates' identifications had failed (incomplete or lack of identification) or were discordant between Dakar and Marseille, the isolates were systematically tested with an appropriate molecular biology method based on the sequencing of 16S rRNA, 18S rRNA, or *rpoB* sequences. In addition, for the *Bacillus cereus* group, additional phenotypic tests were performed to differentiate *Bacillus cereus* from *Bacillus anthracis* and *Bacillus thuringiensis*, as previously reported [9]. All the isolates were transported from Senegal to France in Portagerm Amies Agar swab transport tubes (bioMérieux) at room temperature.

MALDI-TOF MS analysis in Dakar

MALDI-TOF MS analysis was performed in Dakar with a VITEK® MS-RUO version 1.0 (bioMérieux). Each isolated colony was picked and placed in a single well of a disposable, barcode-labeled target slide (VITEK MS-DS, bioMérieux) using a 1-µl plastic loop. Then, each colony was covered with 1.0 µl of a saturated solution of alpha-cyano-4-hydroxycinnamic acid matrix (VITEK MS-CHCA, bioMérieux) and air-dried. Two spots were systematically created for each colony. For each assay, a strain of *Escherichia coli* (Lyfocults *E. coli* ATCC # 8739, bioMérieux) was also analyzed for quality control, including the instrument calibration and the positive controls.

Analyses of the obtained spectra were performed using the Saramis database version 4.0 (bioMérieux). The Saramis software color-codes identification results (by default) according to confidence levels as follows: 99.9%, dark green; 99.8% to 90.0%, light green; 89.9% to 85.0%, yellow; and 84.9% to 70.0%, white. Confidence levels between 70.0% and 99.9% were considered correct identification at the genus and species levels.

MALDI-TOF MS analysis in Marseille

MALDI-TOF MS analysis was performed with a MicroFlex LT mass spectrometer (Bruker Daltonik) as reported [5,6,10]. Each isolated colony was deposited on a MALDI-TOF MS target Microflex (Bruker Daltonik) as above. Then, each colony was overlaid with 2 µL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 2.5% tri-fluoracetic-acid), and the matrix-sample was crystallized by air-drying at room temperature, as previously described [5,6,10]. Two spots were systematically created for each colony. For each assay, a strain of *Escherichia coli* (DH5 alpha, Bruker Daltonik) was also analyzed for quality control.

The analyses of the obtained spectra were performed using our personal database [5,6], the Bruker database updated with a laboratory collection of spectra from clinical isolates identified using molecular sequencing (primarily 16S rRNA sequencing) [6].

The criteria for identification were previously reported [5]. An isolate was considered correctly identified by MALDI-TOF MS if both spectra had a score $\geq$1.9 for species identification or $\geq$1.7 for genus identification [5].

Finally, for both MALDI-TOF MS systems replicates and positive controls are used in a routine manner, the *E. coli* positive controls must be correctly identified and a same identification applying the score of each software package must be observed for the replicate spots of the analyzed colony in order to conclude to the identification of the microorganism.

Identification by molecular biology methods

Molecular analyses using PCR and sequencing were performed targeting 16S ribosomal RNA (rRNA), 18S rRNA, or *rpoB* sequences, as previously described [11–14]. DNA from isolates was extracted using the MagNA Pure LC kit DNA isolation kit III with the EZ-1 biorobot (Qiagen, Hilden, Germany) according to the manufacturer's instructions. All the primers used for PCR and the sequencing of 16S rRNA, 18S rRNA, and *rpoB* targets are summarized in Table 1 [11–17]. PCR products were purified using the PCR kit Nucleofast 96 (Macherey-Nagel, Hoerdt, France), and sequencing was performed with the Big Dye Terminator, version 1.1 sequencing kit (Perkin-Elmer, Coignieres, France) according to the manufacturers' instructions. Products of the sequencing reaction were purified, and sequences were analyzed on an ABI PRISM 3130X Genetic Analyzer (Applied Biosystems, California, USA) [11,11,18,19]. The sequences were assembled and amended with CodonCode Aligner v4.1.1 software (CodonCode Corporation, USA). Then, a correct consensus sequence was saved and

compared with the GenBank database using the BLAST software (http://blast.ncbi.nlm.nih.gov/Blast.cgi). An isolate was correctly identified when it yielded > 98.7% sequence identity for the 16S rRNA sequence and > 97% sequence identity for the *rpoB* sequence with the closest bacterial species sequence in GenBank [11,13,14,20–22].

RESULTS

Overall, 191 out of 202 clinical isolates from Senegal were included in the analyses; 11 isolates were excluded from the analysis due to either contamination or an inability to be cultured in France. False-positive identification at the species level was observed for 6.3% (11/175) of identified isolates in Dakar and for 4.8% (9/184) in Marseille. No false-positive identification was observed at the genus level in Marseille and Dakar.

Concordant identification between the 2 MALDI-TOF MSs at the species level

Overall, 153 out of 191 isolates (80.1%) yielded an accurate and concordant identification at the species level with the 2 different MALDI-TOF MSs in Dakar and Marseille. The 153 isolates correctly identified at the species level by the two MALDI-TOF MSs were from 4 phyla (Table 2), with 63 *Firmicutes* (40.8%), 83 *Proteobacteria* (54.6%), 3 *Actinobacteria* (2%), and 4 *fungi* (2.6%). Among the 63 *Firmicutes*, 1 *Aerococcus viridans;* 2 *Bacillus megaterium*; 7 isolates of *Enterococcus,* including *Enterococcus faecalis* and *Enterococcus faecium*; 36 *Staphylococcus*; and 16 isolates of *Streptococcus*, including 11 *Streptococcus agalactiae,* were properly identified.

Among the 83 *Proteobacteria,* 56 *Enterobacteria,* 26 non-fermentative Gram-negative bacteria, and 1 *Haemophilus influenzae* were correctly identified. Among the 56 *Enterobacteria*, 14 *Enterobacter cloacae*, 10 *Klebsiella pneumoniae*, 10 *Escherichia*

coli, 5 *Morganella morganii*, 4 *Proteus mirabilis*, 3 *Salmonella enterica*, 2 *Citrobacter koseri,* 2 *Enterobacter georgoviae*, 2 *Klebsiella oxytoca,* 1 *Serratia marcescens*, and 1 *Kluyvera ascorbata* were correctly identified. Among the 26 non-fermentative Gram-negative bacteria, 11 *Pseudomonas aeruginosa* and 9 *Acinetobacter baumannii* were properly identified. Among the 3 *Actinobacteria*, 2 *Corynebacterium amycolatum* and 1 *Corynebacterium striatum* were correctly identified. Four fungi, including 3 *Candida albicans* and 1 *Candida tropicalis*, were also properly identified.

Concordant identification between the 2 MALDI-TOF MSs at only the genus level

Overall, 174 out of 191 isolates (91.1%) yielded an accurate and concordant identification at the genus level. Among them, 21 isolates were only concordant at the genus level but not at the species level (Table 3). One isolate of oral streptococci (*Streptococcus oralis*) was not properly identified with either of the MALDI-TOF MSs.

Five isolates were correctly identified with the MALDI-TOF MS in Dakar but not with the MS in Marseille: 2 isolates of *Bacillus cereus*, 1 *Enterobacter gergoviae*, 1 *Salmonella enterica*, and 1 *Staphyloccus warneri*.

Fifteen isolates were properly identified with the MALDI-TOF MS in Marseille but not with the MS in Dakar. Among these, 5 isolates of *Achromobacter xylosoxidans* were systematically inaccurately identified as *Achromobacter denitrificans*. Three isolates of *E. cloacae* was only identified as *Enterobacter* sp. One isolate of *Burkholderia cepacia* and 1 isolate of *Acinetobacter baumannii* were also only identified at the genus level. One isolate of *Corynebacterium striatum* was incorrectly identified as *C. amycolatum*, and 1 isolate of *Streptococcus constellatus* was misidentified as *S. anginosus*. Finally, 1 isolate of *S. haemolyticus* and 1 isolate of S. *simulans* were also inaccurately identified (Table 3).

Lack of identification by the MALDI-TOF MS in Dakar or in Marseille

One isolate was accurately identified as *Escherichia hermannii* in Dakar but not in Marseille; PCR and sequencing of the *rpoB* sequence confirmed the identification of *Escherichia hermannii*. Sixteen out of 191 (8.4%) isolates were not identified by MALDI-TOF MS in Dakar (Figure 1). Among these 16 isolates, 6 were correctly identified at the species level in Marseille (2 isolates of *Exiguobacterium aurantiacum*, 1 isolate of *A. baumannii,* 1 of *Kytococcus schroeteri,* 1 of *Bacillus flexus,* and 1 of *C. albicans*). Four isolates were only identified at the genus level in Marseille, and molecular biology with sequencing was necessary to allow their identification (1 isolate of *Corynebacterium aurimucosum,* 1 of *S. haemolyticus,* 1 of *Paenibacillus amylolyticus,* and 1 of *B. cereus*). Finally, 6 (3%) isolates were not identified with either MALDI-TOF MS. PCR followed by sequencing allowed the identification of 1 isolate of *Rothia mucilaginosa,* 1 of *Staphylococcus arlettae,* 1 of *Bacillus amyloliquefaciens,* 1 of *Bacillus nealsonii,* and 1 of *Exiguobacterium profundum*. In addition, the last isolate exhibited 95% 16S rRNA nucleotide sequence identity with *Necropsobacter rosorum* (NR_114550.1), the phylogenetically closest validated species (Figure 1), suggesting that this isolate corresponded to a new bacterial species, which we named "*Necropsobacter massiliensis*". The Genbank accession number for the 16S rRNA sequence of "*Necropsobacter massiliensis*" is HG428679. The full genome sequencing and the characterization of "*Necropsobacter massiliensis*" are ongoing.

DISCUSSION

MALDI-TOF MS has become a revolutionary technique in the world of clinical microbiology, allowing the quick and accurate identification of clinical pathogenic microorganisms, including bacteria and fungi [3,5,23]. Thus, this newly developed diagnostic tool is increasingly being employed in clinical microbiology laboratories.

Because microbe identification is sometimes difficult in Africa, we have tested a MALDI-TOF MS in Dakar, Senegal. The impact of the installation of this MALDI-TOF MS was so great that other phenotypic-based identification was immediately stopped. Subsequently, the data obtained in this study have confirmed the use of MALDI-TOF MS to hospital in Dakar for microbial identification. Indeed, in total, 80.1% and 91% of isolates were accurately identified at the species and genus levels, respectively, using MALDI-TOF MS in Dakar. These data correspond well to previous work performed in our laboratory in Marseille, France (84.1% of 1,660 tested isolates accurately identified at the species level) [5].

E. coli, *K. pneumoniae*, *S. agalactiae*, *A. baumannii*, *P. aeruginosa*, *S. aureus*, *S. haemolyticus*, *E. cloacae*, *E. faecalis*, and *S. epidermidis* are the 10 most commonly identified bacteria in the laboratory of the Hôpital Principal de Dakar (personal data from the hospital), representing 94.2% of all bacteria routinely identified in the laboratory. All these bacteria were accurately identified with the MALDI-TOF MS in Dakar. Thus, MALDI-TOF MS is a powerful tool for routine bacterial identification in Africa because it allows for the rapid identification of the species most frequently observed there.

The most frequent misidentification observed in Dakar was at the species level for *Achromobacter xylosoxidans*, which was systematically misidentified as *Achromobacter denitrificans* with the Saramis software [24]. Rare misidentifications of other non-fermentative Gram-negative bacteria, such as *Acinetobacter* spp. and *Burkholderia* spp., were also observed. Overall, these misidentifications were comparable to those in other studies that evaluated MALDI-TOF MS performance for the identification of non-fermentative Gram-negative bacteria [24]. Among the *Enterobacteria*, 3 isolates of *E. cloacae* were identified only at the genus level in Dakar. These data are surprising

because 16 other isolates of *E. cloacae* were properly identified. This discrepancy may be explained by the fact that all these 3 isolates produced extended-spectrum beta-lactamases. Indeed, it seems that the Saramis database version 4.0 sometimes has difficulties identifying extended-spectrum beta-lactamase-producing *Enterobacter* at the species level; this was not reported using the Vitek MS *in vitro* diagnostic (IVD) system with the IVD database version 2.0 and MYLA software [25]. Poor identifications are also sometimes associated with technically poor preparation linked to the quality of deposits and impurities of the deposited colony [2]. New deposits might allow an accurate identification.

Coryneform bacteria include a diverse range of bacterial genera grouped together as aerobic non-spore-forming Gram-positive bacilli. Of these, the genus *Exiguobacterium* was first isolated from potato-processing effluent in 1983 and identified as *Exiguobacterium aurantiacum* [26]. *E. aurantiacum* has only been recently reported as a human pathogen with six isolates obtained from patients with bacteremia [26]. Its lack of identification in Dakar was due to its absence from the database. Another bacterium from the genus *Exiguobacterium*, *Exiguobacterium profundum*, was not identified by either MALDI-TOF. *Exiguobacterium profundum* is a recently discovered bacterial species that was first described in 2007 [27]. It was first reported in humans in 2013 with an isolate obtained from a patient with bacteremia and identified using 16S rRNA sequencing [28]. *Exiguobacterium profundum* was also absent from the database.

Several misidentifications or non-identifications of *Bacillus* species were also observed in Marseille and Dakar. The main limit was observed among the *Bacillus cereus* group [29,30]. These identification difficulties may be linked to the fact that these bacteria have very similar spectral profiles associated with the presence of

common protein biomarkers [29]. In addition, on the basis of genetic evidence, it has been proposed that *B. anthracis*, *B. cereus*, and *B. thuringiensis* belong to the same species (*B. cereus* sensu lato), but the status of separate species has been retained due to their remarkably different virulence [31]. Neither MALDI-TOF MS was able to identify *Bacillus nealsonii*, and *Bacillus flexus* was only identified in Marseille. These identification problems were linked to the absence of spectra for these bacteria in the databases. In addition, both MALDI-TOF MSs failed twice to identify oral *Streptococci* accurately at the species level; these difficulties have been reported previously [1,3,5,6,32]. Finally, on the basis of the data observed in Dakar during this study (showing for example that *A. xylosoxidans* has been systematically falsely identified as *A. denitrificans*) and those of the literature which help to better interpret the data, we speculate that no more than 3% of false-positive species identifications might be detected in the real-life application of MALDI-TOS MS in Dakar (Reference).

The improvement of microbial identification is based on the regular addition of new spectra to the databases. For example, in our study, an isolate of *K. schroeteri* was not identified in Dakar, whereas it was identified in Marseille. In fact, an isolate of *K. schroeteri* was first cultivated from a valvular biopsy of a patient with endocarditis in our hospital [33]. This isolate was identified using PCR and 16S rRNA sequencing [33], and its spectrum was added to our database. Subsequently, other isolates of *K. schroeteri* were obtained in a human gut microbiome study performed by our team, and their spectra have been systematically added in our database [34]. Overall, MALDI-TOF MS databases are expanding at a rapid rate, allowing the improvement of the identification but also limiting the value of comparisons of the current results with those of older studies. Finally, 1 of the isolates unidentified by the 2 MALDI-TOF MSs corresponds to a new species of the genus *Necropsobacter* and the *Pasteurellaceae*

family. It was named *Necropsobacter massiliensis*, and its full genome sequence is currently being annotated.

Currently, 2 different MALDI-TOF MS are mainly available. The mass spectrometer of bioMérieux is a large floor model whereas that of Bruker is a desktop instrument. Each system includes spot target plates but the plates are reusable steel targets for Bruker whereas those of bioMérieux are disposable plastic targets. More recently, an extensive comparison was performed on the 2 available MALDI-TOF MS platforms in USA with the most current database versions (Vitek MS and Bruker Microflex LT MALDI-TOF MS). The 2 systems were analytically comparable, with overall isolate identification well above 90% for Food and Drug Administration (FDA)-approved and non-FDA-approved species. Both instruments received high ratings in the user evaluation. However, 75% of the evaluating technologists favored the Viteck MS in a qualitative user assessment.

The robust microbe identification by MALDI-TOF MS in both Dakar and Marseille demonstrates that MALDI-TOF MS can be used as a first-line tool in clinical microbiology laboratories world-wide.

292 **Figure legends.**

293 **Figure 1.** Sixteen clinical isolates not identified in Dakar (Senegal).

Acknowledgements.

We thank the staff of the Hôpital Principal de Dakar (Senegal), especially Dr. Elimane Mbaye and Mr. Diene Bane, for their contributions.

Conceived and designed the experiments: BF BW DR FF. Performed the experiments: CIL BF BSB SD MWG. Analyzed the data: CIL BF BSB OM NF BW DR FF. Contributed reagents/materials/analysis tools: BF BSB SD MWG OM CS NF YD. Wrote the paper: CIL BF BW DR FF.

Table 1. List of all primers used for PCR and sequencing analyses in this study.

Targeted sequences	Primers	Sequences (5' – 3')	Method	References
16S rRNA	Fd1	AGA GTT TGA TCC TGG CTC AG	PCR	[19]
	Rp2	ACG GCT ACC TTG TTA CGA CTT	PCR	[19]
	536F	CAG CAG CCG CGG TAA TAC	Sequencing	[19]
	536R	GTA TTA CCG CGG CTG CTG	Sequencing	[19]
	800F	ATT AGA TAC CCT GGT AG	Sequencing	[19]
	800R	CTA CCA GGG TAT CTA AT	Sequencing	[19]
	1050F	TGT CGT CAG CTC GTG	Sequencing	[19]
	1050R	CAC GAG CTG ACG ACA	Sequencing	[19]
18S rRNA	NS5 (F)	AAC TTA AAG GAA TTG ACG GAA G	PCR and Sequencing	[15]
	NS6 (R)	GCA TCA CAG ACC TGT TAT TGC CTC	PCR and Sequencing	[15]
Corynebacteria rpoB	C2700F	CGW ATG AAC ATY GGB CAG GT	PCR and Sequencing	[16]
	C1330R	TCC ATY TCR CCR AAR CGC TG	PCR and Sequencing	[16]
Enterobacteria rpoB	CM7 (F)	AAC CAG TTC CGC GTT GGC CTG G	PCR and Sequencing	[17]
	CM31b (R)	CCT GAA CAA CAC GCT CGG A	PCR and Sequencing	[17]
Staphylococci rpoB	Staph_F	AAC CAA TTC CGT ATI GGT TT	PCR and Sequencing	[18]
	Staph_R	CCG TCC CAA GTC ATG AAA C	PCR and Sequencing	[18]
Streptococci rpoB	Strepto_F	AAR YTI GGM CCT GAA GAA AT	PCR and Sequencing	[12]
	Strepto_R	TGI ART TTR TCA TCA AAC ATG TG	PCR and Sequencing	[12]

Table 2. 153 clinical isolates correctly identified at the species level by MALDI-TOF MS, first in Dakar (Senegal) and then in Marseille (France).

Phylum	Genus	Species	No. of isolates
Actinobacteria	*Corynebacterium*	*Corynebacterium amycolatum*	2
		Corynebacterium striatum	1
Firmicutes	*Aerococcus*	*Aerococcus viridans*	1
	Bacillus	*Bacillus megaterium*	2
	Enterococcus	*Enterococcus faecalis*	6
		Enterococcus faecium	1
	Staphylococcus	*Staphylococcus arlettae*	1
		Staphylococcus aureus	9
		Staphylococcus capitis	1
		Staphylococcus caprae	1
		Staphylococcus cohnii	4
		Staphylococcus epidermidis	1
		Staphylococcus haemolyticus	9
		Staphylococcus hominis	4
		Staphylococcus saprophyticus	2
		Staphylococcus simulans	2
		Staphylococcus warneri	2
	Streptococcus	*Streptococcus agalactiae*	11
		Streptococcus anginosus	1
		Streptococcus dysgalactiae	1
		Streptococcus pyogenes	3
		Streptococcus salivarius	1
Proteobacteria	*Acinetobacter*	*Acinetobacter baumannii*	9
		Acinetobacter lwoffii	1
		Acinetobacter radioresistens	3
	Citrobacter	*Citrobacter koseri*	2
	Enterobacter	*Enterobacter cloacae*	14
		Enterobacter gergoviae	2
	Escherichia	*Escherichia coli*	10
	Klebsiella	*Klebsiella oxytoca*	2
		Klebsiella pneumoniae	10
	Kluyvera	*Kluyvera ascorbata*	1
	Haemophilus	*Haemophilus influenzae*	1
	Morganella	*Morganella morganii*	5
	Proteus	*Proteus mirabilis*	4
		Proteus penneri/vulgaris	2
	Pseudomonas	*Pseudomonas aeruginosa*	11
		Pseudomonas putida	1
	Salmonella	*Salmonella enterica*	3
	Serratia	*Serratia marcescens*	1
	Stenotrophomonas	*Stenotrophomonas maltophilia*	1
Fungi	*Candida*	*Candida albicans*	3
		Candida tropicalis	1
Total			**153**

Table 3. Twenty-one MALDI-TOF MS identifications with concordance only at the genus level in Dakar (Senegal) and Marseille (France); Species identification was resolved by molecular biology analyses coupled with phenotypic tests for the *Bacillus cereus* group.

MALDI-TOF MS, Dakar (No.)	MALDI-TOF MS, Marseille (No.)	Identification using molecular biology tools
Achromobacter denitrificans (5)	*Achromobacter xylosoxidans* (5)	*A. xylosoxidans*[1] (5) [HQ288926: 98.8% and 99.3% ; GQ889256: 99.6%; AF411020: 100% and 99.9%]*
Acinetobacter baumannii	*Acinetobacter junii*	*A. junii*[1] [HE651919: 99%]
Corynebacterium amycolatum	*Corynebacterium striatum*	*C. striatum*[2] [HE586297: 99.7%]
Staphylococcus aureus	*Staphylococcus haemolyticus*	*S. haemolyticus*[1] [NR_074994: 99.3%]
Staphylococcus haemolyticus	*Staphylococcus simulans*	*S. simulans*[1] [KC849411: 97.9%]
Streptococcus anginosus	*Streptococcus constellatus*	*S. constellatus*[4] [AF535184: 98.2%]
Acinetobacter sp.	*A. baumannii*	*A. baumannii*[1] [EU734813: 97.8%]
Enterobacter sp. (3)	*Enterobacter cloacae* (3)	*E. cloacae*[3] (3) [JQ435865: 97.9% ; 99.1% and 98.1%]
Burkholderia sp.	*Burkholderia cepacia*	*B. cepacia*[1] [EU742139: 99.2%]
Streptococcus gordonii	*Streptococcus mitis*	*Streptococcus oralis*[4] [AF535168: 98%]
Bacillus cereus	*Bacillus anthracis*	*B. cereus* group[1,§] [CP003187: 97.9%]
Bacillus cereus	*Bacillus* sp.	*B. cereus* group[1,§] [CP003187: 99.3%]
Enterobacter georgoviae	*Enterobacter asburiae*	*E. georgoviae*[3] [AJ566945: 99%]
Salmonella enterica	*Salmonella typhi*	*S. enterica* subsp. *enterica* serovar paratyphi A[3] [JQ728878: 100%]
Staphylococcus warneri	*Staphylococcus pasteuri*	*S. warneri*[1] [HQ407248: 99%]

No: number of isolates, if greater than 1

[1] 16S rRNA sequencing

[2] *rpoB* gene sequencing for *Corynebacteria*

[3] *rpoB* gene sequencing for *Enterobacteria*

[4] *rpoB* gene sequencing for *Streptococci*

[*] The GenBank accession numbers and the percentage of homology obtained for all the sequenced bacteria are given between the brackets.

[§] The isolates were both hemolytic on Columbia blood agar, and no parasporal crystals were observed in sporulated cultures, leading to the identification of *Bacillus cereus*.

References.

1. Clark AE, Kaleta EJ, Arora A, Wolk DM (2013) Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev 26: 547-603.

2. Patel R (2013) Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry in Clinical Microbiology. Clin Infect Dis 57: 564-572.

3. Bizzini A, Durussel C, Bille J, Greub G, Prod'hom G (2010) Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. J Clin Microbiol 48: 1549-1554.

4. Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, Francois P, Schrenzel J (2010) Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. J Clin Microbiol 48: 1169-1175.

5. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D (2009) Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 49: 543-551.

6. Seng P, Abat C, Rolain JM, Colson P, Gouriet F, Fournier PE, Drancourt M, La Scola B, Raoult D (2013) Emergence of rare pathogenic bacteria in a clinical microbiology laboratory: impact of MALDI-TOF mass spectrometry. J Clin Microbiol 51: 2182-2194.

7. van Veen SQ, Claas EC, Kuijper EJ (2010) High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight

mass spectrometry in conventional medical microbiology laboratories. J Clin Microbiol 48: 900-907.

8. Fall B, Lo CI, Samb-Ba B, Perrot N, Diawara S et al. (2015) The Ongoing Revolution of MALDI-TOF Mass Spectrometry for Microbiology Reaches Tropical Africa. Am J Trop Med Hyg 92:641-647.

9. Logan NA, Popovic T, Hoffmaster A (2007) *Bacillus* and other aerobic endospore-forming bacteria. In: Murray PR, Baron EJ, Jorgensen JH, Landry ML, Pfaller MA, editors. Manual of Clinical Microbiology 9th edition. Washington, D.C.. pp. 455-473.

10. Dridi B, Raoult D, Drancourt M (2012) Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of Archaea: towards the universal identification of living organisms. APMIS 120: 85-91.

11. Drancourt M, Berger P, Raoult D (2004) Systematic 16S rRNA gene sequencing of atypical clinical isolates indentified 27 new bacterial species associated with humans. J Clin Microbiol 42: 2197-2202.

12. Drancourt M, Roux V, Fournier PE, Raoult D (2004) rpoB gene sequence-based identification of aerobic Gram-positive cocci of the genera *Streptococcus*, *Enterococcus*, *Gemella*, *Abiotrophia*, and *Granulicatella*. J Clin Microbiol 42: 497-504.

13. Fenollar F, Roux V, Stein A, Drancourt M, Raoult D (2006) Usefulness of 16S rDNA PCR and sequencing in bone and joint infections. An experience on 525 samples. J Clin Microbiol 44: 1018-1028.

14. Khamis A, Raoult D, La Scola B (2004) rpoB gene sequencing for identification of *Corynebacterium* species. J Clin Microbiol 42: 3925-3931.

15. Gutzmer R, Mommert S, Kuttler U, Werfel T, Kapp A (2004) Rapid identification and differentiation of fungal DNA in dermatological specimens by LightCycler PCR. J Med Microbiol 53: 1207-1214.

16. Suzuki T, Iihara H, Uno T, Hara Y, Ohkusu K, Hata H, Shudo M, Ohashi Y (2007) Suture-related keratitis caused by Corynebacterium macginleyi. J Clin Microbiol 45: 3833-3836.

17. Drancourt M, Bollet C, Carta A, Rousselier P (2001) Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. Int J Syst Evol Microbiol 51: 925-932.

18. Drancourt M, Raoult D (2002) rpoB gene sequence-based identification of *Staphylococcus* species. J Clin Microbiol 40: 1333-1338.

19. Al Masalma M, Armougom F, Scheld WM, Dufour H, Roche PH, Drancourt M, Raoult D (2009) The expansion of the microbiological spectrum of brain abscesses with use of multiple 16S ribosomal DNA sequencing. Clin Infect Dis 48: 1169-1178.

20. Adekambi T, Drancourt M, Raoult D (2009) The rpoB gene as a tool for clinical microbiologists. Trends Microbiol 17: 37-45.

21. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kampfer P, Maiden MC, Nesme X, Rossello-Mora R, Swings J, Truper HG, Vauterin L, Ward AC, Whitman WB (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52: 1043-1047.

22. Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173: 697-703.

23. Ling H, Yuan Z, Shen J, Wang Z, Xua Y (2014) Accuracy of MALDI-TOF mass spectrometry for the identification of clinical pathogenic fungi: a meta-analysis. J Clin Microbiol. 52:2573-2582.

24. Manji R, Bythrow M, Branda JA, Burnham CA, Ferraro MJ, Garner OB, Jennemann R, Lewinski MA, Mochon AB, Procop GW, Richter SS, Rychert JA, Sercia L, Westblade LF, Ginocchio CC (2014) Multi-center evaluation of the VITEK(R) MS system for mass spectrometric identification of non-Enterobacteriaceae Gram-negative bacilli. Eur J Clin Microbiol Infect Dis 33: 337-346.

25. Richter SS, Sercia L, Branda JA, Burnham CA, Bythrow M, Ferraro MJ, Garner OB, Ginocchio CC, Jennemann R, Lewinski MA, Manji R, Mochon AB, Rychert JA, Westblade LF, Procop GW (2013) Identification of Enterobacteriaceae by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using the VITEK MS system. Eur J Clin Microbiol Infect Dis 32: 1571-1578.

26. Pitt TL, Malnick H, Shah J, Chattaway MA, Keys CJ, Cooke FJ, Shah HN (2007) Characterisation of *Exiguobacterium aurantiacum* isolates from blood cultures of six patients. Clin Microbiol Infect 13: 946-948.

27. Crapart S, Fardeau ML, Cayol JL, Thomas P, Sery C, Ollivier B, Combet-Blanc Y (2007) *Exiguobacterium profundum* sp. nov., a moderately thermophilic, lactic acid-producing bacterium isolated from a deep-sea hydrothermal vent. Int J Syst Evol Microbiol 57: 287-292.

28. Cheng A, Liu CY, Tsai HY, Hsu MS, Yang CJ, Huang YT, Liao CH, Hsueh PR (2013) Bacteremia caused by *Pantoea agglomerans* at a medical center in Taiwan, 2000-2010. J Microbiol Immunol Infect 46: 187-194.

29. Castanha ER, Fox A, Fox KF (2006) Rapid discrimination of *Bacillus anthracis* from other members of the B. cereus group by mass and sequence of "intact" small acid soluble proteins (SASPs) using mass spectrometry. J Microbiol Methods 67: 230-240.

30. Ryzhov V, Hathout Y, Fenselau C (2000) Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Appl Environ Microbiol 66: 3828-3834.

31. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto AB (2000) *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*-one species on the basis of genetic evidence. Appl Environ Microbiol 66: 2627-2630.

32. Tan KE, Ellis BC, Lee R, Stamper PD, Zhang SX, Carroll KC (2012) Prospective evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: a bench-by-bench study for assessing the impact on time to identification and cost-effectiveness. J Clin Microbiol 50: 3301-3308.

33. Renvoise A, Roux V, Casalta JP, Thuny F, Riberi A (2008) *Kytococcus schroeteri*, a rare agent of endocarditis. Int J Infect Dis 12: 223-227.

34. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape JF, Koonin EV, La Scola B,

Raoult D (2012) Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 18:1185-93.

Article 3:

« The ongoing revolution of MALDI-TOF mass spectrometry for microbiology reaches tropical Africa»

Bécaye Fall[1], **Cheikh Ibrahima Lo**[2], Bissoume Samb-Ba[1], Nadine Perrot[3], Silman Diawara[1], Mamadou Wague Gueye[1], Kowry Sow[1], Maxence Aubadie-Ladrix[2], Oleg Mediannikov[2], Cheikh Sokhna[2], Yaya Diemé[1], Sonia Chatellier, Boubacar Wade[1], Didier Raoult[2], and Florence Fenollar[2]

[1]Hôpital Principal, Dakar, Senegal

[2]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Aix-Marseille Université, Marseille, France and Dakar, Senegal

[3]bioMérieux, La Balme-Les-Grottes, France

Publié dans:

The American Journal of Tropical Medicine and Hygiene

March 2015 Volume 92, Issue 3, pages 641-647.

http://www.ncbi.nlm.nih.gov/pubmed/25601995

Résumé de l'article 3: ***«The ongoing revolution of MALDI-TOF mass spectrometry for microbiology reaches tropical Africa»***

Contexte:

La spectrométrie de masse MALDI-TOF représente une révolution dans l'identification de routine des agents pathogènes rencontrés dans les laboratoires de microbiologie clinique. En effet, un appareil MALDI-TOF a été installé en Afrique tropicale pour la première fois dans le laboratoire clinique de microbiologie de l'Hôpital Principal de Dakar (Sénégal) (34) dans le but de lutter contre les maladies infectieuses tropicales et émergentes en détectant de nouveaux microorganismes (bactéries ou champignons).

Méthode:

Le coût actuel d'un appareil MALDI-TOF varie entre 100 000 à 200 000 euros. Cette contrainte financière retarde son acquisition dans les pays en développement. Le Sénégal est doté d'un MALDI-TOF (Vitek® MS RUO, bioMérieux, Marcy l'Etoile, France) grâce aux soutiens financiers de ses partenaires français comme l'IRD, la Méditerranéen Infection et l'ambassade de la France au Sénégal. Une formation technique de quatre personnes dont moi-même a permis l'utilisation autonome de l'appareil pour le diagnostic de routine et la recherche.

Des souches fraiches sont isolées à partir de 2640 prélèvements cliniques chez des patients consultants ou hospitalisés à l'Hôpital Principal de Dakar (HPD), pendant une période de dix mois. Toutes ces souches sont identifiées par MALDI-TOF. Par ailleurs, pour évaluer la performance du MALDI-TOF Vitek® MS (bioMérieux) installé à Dakar, 93 souches dont celles fréquemment isolées dans les laboratoires de microbiologie clinique et préalablement testées avec le MALDI-TOF de Dakar, sont par la suite envoyées à Marseille pour confirmer leur identification avec le MALDI-TOF Microflex LT (Bruker Daltonics).

Le principe d'identification par MALDI-TOF repose sur la comparaison entre les spectres obtenus à partir d'une analyse des protéines de l'échantillon testé et les spectres existant dans la base de données du MALDI-TOF. Le référentiel d'identification varie selon le fabricant de l'appareil. Ainsi dans le cas du MALDI-TOF Vitek MS installé à Dakar dont le constructeur est la société bioMérieux, les identifications dépendent d'un logiciel Saramis qui affiche un code de couleur par défaut avec un degré de confiance qui varie entre 60 à 99.9. Ainsi une correspondance parfaite entre le spectre obtenu et le spectre unique du germe ou groupe de germes étudiés donne un degré de confiance de 99.9. Par contre, quand le degré de confiance est inférieur à 60, l'organisme est considéré comme non identifié par le système. Enfin si le degré de confiance est supérieur à 60 et inférieur à 99, l'identification n'est pas précise et des tests de confirmation seront effectués.

Résultats:

Dans l'ensemble, 2 429 microorganismes isolés chez les patients sont soumis à l'identification par MALDI-TOF. Parmi ceux-ci 2 082 bactéries (85,7%) et 206 champignons (8,5%) sont correctement identifiés au niveau de l'espèce. De même 109 bactéries (4,5%) et 16 autres bactéries (0,75%) étaient respectivement identifiées jusqu'au niveau du genre et de la famille. Enfin il y a 16 autres souches (0,75%) qui n'ont pas été identifiées par MALDI-TOF.

Par ailleurs, nous avons noté lors des identifications, une forte prévalence d'*Escherichia coli* (25,8%) suivie de celles de *Klebsiella pneumoniae* (14,8%), *Streptococcus agalactiae* (6,2%), *Acinetobacter baumannii* (6,1%), *Pseudomonas aeruginosa* (5,9%), et *Staphylococcus aureus* (5,9%). Le MALDI-TOF a permis aussi de détecter des bactéries et des champignons rares.

Parmi les 206 champignons identifiés, 197 étaient du genre *Candida* (95,6%) dont l'espèce *Candida albicans* (47,6%) a été identifiée plus de cinquante fois.

La majorité des 109 bactéries identifiées seulement au niveau genre, correspondait au genre *Streptococcus* (44%). Les autres mal identifiées concernaient principalement le phylum des *Proteobacteria.* Du fait des similarités de protéines ribosomales entre les espèces *Escherichia coli* et *Shigella*, le MALDI-TOF ne parvenait pas à distinguer correctement ces espèces appartenant à la famille des *Enterobacteriaceae.*

Conclusion:

En conclusion, la spectrométrie de masse MALDI-TOF offre une identification rapide et précise des bactéries d'intérêt clinique telles que celles à gram positif ou négatif ainsi que celles anaérobiques fastidieuses, détectées en microbiologie de routine. Les résultats obtenus prouvent que le MALDI-TOF est un puissant outil d'identification des bactéries et champignons avec 94,2% d'identification précise jusqu'au niveau de l'espèce pour les isolats testés dans le laboratoire clinique en Afrique tropicale soient 2 289/2 429 souches détectées.

Le SM MALDI-TOF a permis aussi la détection des microorganismes rares comme les bactéries telles que *Arthrobacter cumminsii* et *Streptococcus australis*, ainsi que des champignons comme *Candida africana* et *Candida nivariensis.* Ceci montre que le MALDI-TOF est un moyen efficace de prévention et de lutte contre les maladies infectieuses fréquentes en Afrique.

Am. J. Trop. Med. Hyg., 92(3), 2015, pp. 641–647
doi:10.4269/ajtmh.14-0406

The Ongoing Revolution of MALDI-TOF Mass Spectrometry for Microbiology Reaches Tropical Africa

Bécaye Fall, Cheikh Ibrahima Lo, Bissoume Samb-Ba, Nadine Perrot, Silman Diawara, Mamadou Wague Gueye, Kowry Sow, Maxence Aubadie-Ladrix, Oleg Mediannikov, Cheikh Sokhna, Yaya Diemé, Sonia Chatellier, Boubacar Wade, Didier Raoult, and Florence Fenollar*

Hôpital Principal, Dakar, Senegal; Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, Research Institute for Development 198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Aix-Marseille Université, Marseille, France; Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, Research Institute for Development 198, Inserm 1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Aix-Marseille Université, Dakar, Senegal; bioMérieux, La Balme-Les-Grottes, France

Abstract. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) represents a revolution in routine pathogen identification in clinical microbiology laboratories. An MALDI-TOF MS was introduced to tropical Africa in the clinical microbiology laboratory of the Hôpital Principal de Dakar (Senegal) and used for routine pathogen identification. Using MS, 2,429 bacteria and fungi isolated from patients were directly assayed, leading to the identification of 2,082 bacteria (85.7%) and 206 fungi (8.5%) at the species level, 109 bacteria (4.5%) at the genus level, and 16 bacteria (0.75%) at the family level. Sixteen isolates remained unidentified (0.75%). *Escherichia coli* was the most prevalent species (25.8%) followed by *Klebsiella pneumoniae* (14.8%), *Streptococcus agalactiae* (6.2%), *Acinetobacter baumannii* (6.1%), *Pseudomonas aeruginosa* (5.9%), and *Staphylococcus aureus* (5.9%). MALDI-TOF MS has also enabled the detection of rare bacteria and fungi. MALDI-TOF MS is a powerful tool for the identification of bacterial and fungal species involved in infectious diseases in tropical Africa.

INTRODUCTION

The routine identification of bacteria and fungi by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) became prevalent 5 years ago and represents a revolution in clinical microbiology laboratories.[1–3] This technique enables the identification of bacteria and fungi in less than 1 hour starting with pure culture without *a priori* knowledge of the types of microorganisms in the sample. MALDI-TOF MS is becoming a powerful tool for routine identification, replacing Gram staining and all fastidious biochemical identifications.[4] The high cost of MALDI-TOF instruments, the limitations of existing bioinformatics tools, and the lack of convenient preparations of the required chemicals previously limited the development of this technology,[5] but more recently, the reduction in the cost of the instrument has facilitated access to this technology.[5] Thus, its use has become widespread in many clinical laboratories, first primarily in Europe and Asia.[1,2,5–11]

Currently, the specific identification of microorganisms in Africa raises several issues, including that there are no regulations governing pathogen identification in many countries. Although culturing capabilities are available in major hospitals in Africa, performance limitations of the biochemical identification methods may still be encountered. When biochemical tests are available, they can still be laborious and difficult to interpret, and they can lead to poor identifications. A series of standardized and miniaturized biochemical tests associated with a database (by numerical identification) can be used to make identification easier and more accurate; however, this approach requires many kits (for *Enterobacteriaceae*, non-*Enterobacteriaceae* Gram-negative bacilli, *Staphylococcus*, *Streptococcus*, and others) along with other dedicated reagents. Kits and reagents should, furthermore, be stored under specific conditions and have expiration dates. Reagent supply issues are frequently associated with potential problems or backorders. Furthermore, biochemical methods are frequently time-consuming, often require knowledge about the type of microorganism being tested, and fail to accurately identify several microorganisms.[1,4,12]

Herein, we implemented MALDI-TOF MS in a clinical microbiology laboratory in an African hospital (Hôpital Principal de Dakar [Principal Hospital of Dakar]) in Dakar, Senegal and evaluated its potential for the reliable and rapid identification of common microorganisms.

MATERIALS AND METHODS

Constraints on the acquisition of MALDI-TOF MS instrumentation in Africa. *Funding for the acquisition of MALDI-TOF MS instrumentation.* The current cost for an MALDI-TOF MS instrument is estimated to be between approximately 100,000 and 200,000 Euros.[5] The acquisition of the MALDI-TOF MS instrument (VITEK MS RUO; bioMérieux, Marcy l'Etoile, France) in use in Dakar was supported by the Méditerranée Infection Hospital–University Institute (IHU Méditerranée Infection; http://www.mediterranee-infection.com/), the Research Institute for Development (IRD; http://www.ird.fr/), and the French Ministry of Foreign Affairs. Created in 2012, the aim of IHU Méditerranée Infection is to fight infectious diseases on a global scale. Since 2007, the Infectiopôle South Foundation, a department of the IHU, promotes north–south trade and the coordination of scientific projects in the field of research on infectious diseases, including financial support for foreign students from the south. IRD is a public French organization involved in research with and for southern countries. Through its research, training,

*Address correspondence to Florence Fenollar, Aix Marseille Université, Unité des Rickettsies, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM 63, CNRS 7278, Research Institute for Development 198, Inserm 1095, 27 Boulevard Jean Moulin, 13005 Marseille, France. E-mail: florence.fenollar@univ-amu.fr

and innovation in partnership, IRD is involved in work in more than 50 countries in Africa, the Mediterranean area, Asia, Latin America, and French overseas territories in other regions.

Cooperation agreement and convention. The Principal Hospital of Dakar (HPD; http://www.hopitalprincipal.sn/) is a public health hospital with special status as a military teaching hospital in Senegal. It has 420 beds and participates in the care of Senegalese patients and other patients from the surrounding area. The Clinical Microbiology Laboratory at the hospital is open 24 hours per day throughout the year. The MALDI-TOF MS was installed in this laboratory. A convention was signed between the HPD, IRD, and IHU Méditerranée Infection. The cooperation agreement stated that the instrument belongs to the HPD. In return, the hospital must ensure the recruitment of staff responsible for operating the MALDI-TOF MS instrument, provide suitable premises for its installation, provide open access to research programs of the IRD, and help other hospitals in the area to identify microorganisms.

Constraints on the installation of the MALDI-TOF MS instrument in Africa. *Technical constraints.* The chemical matrix, which is a unique reagent required for MALDI-TOF MS, must be stored at +4°C when purchased or prepared on the day of use and stored at room temperature. The room containing the MALDI-TOF MS instrument must be thermally isolated, air-conditioned, and protected from insects and dust. Electricity must be supplied continuously. The instrument as well as three computers must each be equipped with inverters in case of voltage dips or brief power outages. To prevent power failure, an electric generator must be provided. The MALDI-TOF MS unit was supplied with several spare parts. Maintenance should be performed one time per year; the first annual maintenance was performed by an engineer of bioMérieux, and it was supported by bioMérieux.

Human constraints. Local personnel operating the MALDI-TOF MS instrument should be specifically trained to use the instrument. Before the initiation of the project, a local operator (B.S.-B.), a PhD student, was trained to prepare target slides, use an MALDI-TOF MS instrument (Bruker Daltonik, Wissembourg, France), and analyze results during a 6-month period in the IHU Méditerranée Infection. In July of 2012, B.S.-B. (recruited then by HPD) and other local operators (B.F., C.I.L., and M.A.-L.) were trained to use the VITEK MS instrument specifically in a 4-day course in Dakar given by two engineers from bioMérieux. After this training course, all were capable of performing MALDI-TOF MS autonomously. A follow-up review of this training was conducted in November of 2012.

Identification of microorganisms using MALDI-TOF MS in Africa. *Bacterial and fungal isolates.* Fresh isolates were obtained from 2,640 specimens in the course of routine clinical work in the HPD clinical laboratory and tested over 10 months during the study period (August of 2012 to May of 2013). All isolates recovered from blood, cerebrospinal fluid, pus, biopsies, the respiratory tract, the urogenital tract, wounds, stool specimens, and devices were prospectively included in the study. The isolates were recovered after the inoculation of clinical specimens on 5% horse blood Mueller–Hinton agar, trypticase soy agar, and MacConkey agar media. Sabouraud agar media was inoculated when required. All media were prepared in the laboratory. In all cases, the cultures were incubated under standard conditions for a minimum of 18 hours at 35–37°C in ambient air with either CO_2 enrichment or in anaerobic atmospheres.

To assess the accuracy of the VITEK MS RUO system for routine bacterial identification, we evaluated 93 strains, including all of the most current detected bacteria in a clinical microbiology laboratory; these bacteria had previously been isolated from patients and identified using a Bruker BioTyper in our laboratory in Marseille, France (Supplemental Table 1). A 100% concordance in identification was observed, allowing us to use the VITEK MS RUO for diagnostic purposes.

MALDI-TOF MS analysis. The isolated colonies were deposited in a single well of a disposable, barcode-labeled target slide (VITEK MS-DS) using a 1.0-μL loop, then overlaid with 1.0 μL of a saturated solution of α-cyano-4-hydroxycinnamic acid matrix (VITEK MS-CHCA; bioMérieux), and air-dried. If the presence of fungi was suspected, 1 μL formic acid solution (VITEK MS-FA; bioMérieux) was first added. Two spots were prepared for each isolated colony. For instrument calibration, an *Escherichia coli* reference strain (Lyfocults *Escherichia coli* ATCC 8739; bioMérieux) was transferred to designated wells on the target slide using the procedure described above. For quality control purposes, positive controls (*E. coli* strains) were analyzed in each assay.

The Biotyper software compared the protein profile of the microorganisms obtained from the Saramis database, version 4.0 (bioMérieux). The Saramis software color-codes identification results (by default) according to confidence levels as follows: 99.9%, dark green; 99.8–90.0%, light green; 89.9–85.0%, yellow; and 84.9–70.0%, white. Identification results obtained between 70.0% and 99.9% confidence were considered to be correct identifications at the genus and species levels.

RESULTS

Overall, 2,429 bacteria and fungi were isolated from 2,640 specimens received in the laboratory (Table 1) and directly tested using MALDI-TOF MS, leading to the identification of 2,082 bacteria (85.7%) and 206 fungi (8.5%) at the species level, 109 bacteria (4.5%) at the genus level, and 16 bacteria (0.75%) at the family level. Sixteen isolates were not identified (16 of 2,429; 0.65%).

Accurate identification at the species level. *Bacterial identification.* Ten bacteria were identified more than 50 times and together, represented 94.2% (1,962 of 2,083) of the bacterial isolates (Figure 1). *E. coli* was the most frequently identified bacterial species (538 of 2,083; 25.8%) followed

TABLE 1
Distribution of 2,640 specimens analyzed at the Principal Hospital, Dakar

Specimens	Number	Percentage
Urine	979	37.1
Pus	471	17.9
Vagina	381	14.4
Blood	289	10.9
Respiratory tract	194	7.4
Stomach	113	4.3
Peripheral devices (catheters, probes, and others)	72	2.7
Feces	61	2.3
Genitalia (other than vagina)	38	1.5
Puncture fluid	13	0.5
Other	29	1

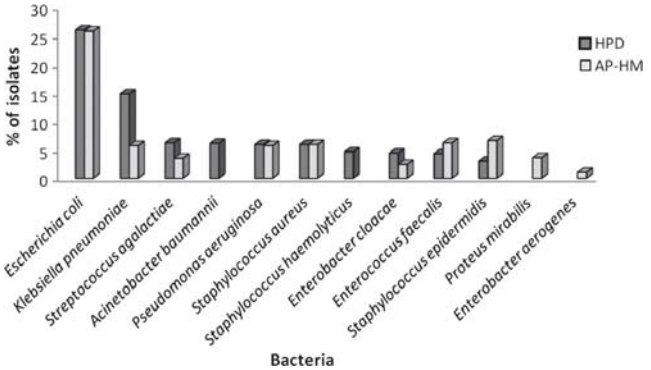


FIGURE 1. Percentage of 10 bacteria most frequently identified using VITEK MALDI-TOF MS in Dakar, Senegal and Marseille, France. AP-HM = Assistance Publique - Hôpitaux de Marseille.

by *Klebsiella pneumoniae* (308; 14.8%), *S. agalactiae* (130; 6.2%), *Acinetobacter baumannii* (128; 6.1%), *Pseudomonas aeruginosa* (124; 5.9%), *S. aureus* (124; 5.9%), *S. haemolyticus* (95; 4.6%), *Enterobacter cloacae* (92; 4.4%), *Enterococcus faecalis* (90; 4.3%), and *S. epidermidis* (63; 3%).

Ten bacteria were identified between 10 and 50 times: *Morganella morganii* (39; 1.9%), *S. hominis* (31; 1.5%), *Proteus mirabilis* (20; 1%), *S. pyogenes* (16; 0.8%), *S. cohnii* (14; 0.7%), *S. saprophyticus* (14; 0.7%), *E. faecium* (13; 0.6%), *Stenotrophomonas maltophilia* (13; 0.6%), *S. anginosus* (13; 0.6%), *S. warneri* (11; 0.5%), and *P. putida* (10; 0.48%).

Among 35 bacteria identified more than 1 time but less than 10 times (Table 2), *Salmonella enterica* (9), *A. radioresistens* (8), *Citrobacter koseri* (8), *Providencia rettgeri* (8), *S. salivarius* (8), *A. junii* (7), *Bacillus cereus* (7), *P. stuartii* (7), *E. asburiae* (6), *K. oxytoca* (6), *P. stutzeri* (6), and *S. parasanguinis* (6) were the most frequent.

Among 34 bacteria identified only one time (Table 3), emerging pathogens, such as *Alloiococcus otitis*, and rare pathogens, such as *Arthrobacter cumminsii* or *S. australis*, were detected.

Fungal identification. Among 206 identified fungi (Table 4), 197 were from the *Candida* genus (95.6%). Only one fungus was identified more than 50 times: *C. albicans* (98; 47.6%). Three fungi were identified between 10 and 50 times: *C. tropicalis* (42; 20.4%), *C. glabrata* (30; 14.6%), and *C. krusei* (14; 6.8%). Five fungi were identified more than 1 time but less than 10 times: *C. parapsilosis* (6), *Aspergillus niger* (3), *C. dubliniensis* (2), *Clavispora lusitaniae* (2), and *Kluyveromyces marxianus* (2). Seven were identified only one time: *A. flavus*, *Microsporum canis*, *Trichosporon asahii*, *Kodamaea ohmeri*, *C. africana* (an emergent and rare pathogen described in 2001 for the first time),[13] *C. nivariensis* (an emergent and rare pathogen described in 2005 for the first time),[14] and *C. utilis* (an industrially important yeast that is rarely reported as a human pathogen, with approximately 10 reported cases found in a PubMed search on May 6, 2014).[15]

Imprecise identification at the genus or family levels. Among 109 bacteria identified at the genus level (Table 5), most were from the genus *Streptococcus* (48 of 109; 44%), including 29 (26.6%) isolates with a misidentification between *S. mitis*, *S. oralis*, and *S. pneumoniae*. The other misidentifications primarily included bacteria from the *Proteobacteria* phylum, with difficulties occurring in identifying bacteria from the *Citrobacter* genus (10; 9.2%), *Achromobacter* genus (10; 9.2%), *Burkholderia* genus (9; 8.3%), and *Aeromonas* genus (7; 6.4%). In 16 cases, bacteria from the *Enterobacteriaceae* family were identified, but MALDI-TOF MS could not accurately distinguish between *E. coli* and *Shigella*, because they are likely pathovars belonging to the same species with similar ribosomal protein patterns.

DISCUSSION

Several studies, including comparative and/or multicenter studies, have already been performed to evaluate and compare the performance of the most diffuse commercial systems of MALDI-TOF MS systems, such as the Bruker BioTyper and the bioMérieux VITEK MS (with both SARAMIS v4.09 and Knowledge Base v2.0 VITEK MS v2.0 systems), by checking the discrepancies through molecular methods and sequencing.[11,16–19] Overall, MALDI-TOF MS identification has been found to be highly accurate for clinically relevant bacteria, including Gram-positive, Gram-negative, and fastidious anaerobic bacteria as well as fungi detected in routine microbiology.[11,16,18,20] Herein, we confirm the power of MALDI-TOF MS in identifying bacteria and fungi, with 94.2% identification accuracy at the species level of isolates tested in a laboratory in tropical Africa (2,289 of 2,430 isolates). In the pioneering work performed in 2010 in our laboratory in Marseille, France, 84.1% of 1,660 tested isolates were accurately identified at the species level, and 11.3% of tested isolates were accurately identified at the genus level.[5] In Dakar, the 10 most commonly identified bacteria also represented 94.2% of all bacteria that were accurately identified at the laboratory. MALDI-TOF MS has also enabled the detection of rare microorganisms, including bacteria, such as *A. cumminsii* and *S. australis*, as well as fungi, such as *C. africana* and *C. nivariensis*.

Most of the misidentifications were caused by the potential inability of MALDI-TOF MS (already highlighted in other

TABLE 2
Thirty-five bacteria identified at the species level from two to nine times

Phylum, genus, and bacterial species	Number of isolates
Actinobacteria	
Corynebacterium	
C. amycolatum	3
C. striatum	5
C. aurimucosum	3
Firmicutes	
Aerococcus	
A. viridans	4
Bacillus	
B. cereus	7
B. pumilus	3
Lactobacillus	
L. delbrueckii	2
Micrococcus	
M. luteus	2
Staphylococcus	
S. auricularis	3
S. lugdunensis	4
S. sciuri	2
S. simulans	2
Streptococcus	
S. dysgalactiae	2
S. parasanguinis	6
S. pneumoniae	5
S. salivarius	8
Proteobacteria	
Achromobacter	
A. xylosoxidans	5
Acinetobacter	
A. junii	7
A. lwoffii	2
A. radioresistens	8
A. schindleri	4
Citrobacter	
C. freundii	3
C. koseri	8
Enterobacter	
E. asburiae	6
E. gergoviae	5
Haemophilus	
H. influenzae	4
Klebsiella	
K. oxytoca	6
Neisseria	
N. meningitidis	2
Plesiomonas	
P. shigelloides	3
Providencia	
P. rettgeri	8
P. stuartii	7
Pseudomonas	
P. fluorescens	2
P. stutzeri	6
Ralstonia	
R. pickettii	2
Salmonella	
S. enterica	9
Serratia	
S. marcescens	4

TABLE 3
Thirty-four bacteria identified at the species level one time

Phylum and genus	Bacterial species
Actinobacteria	
Arthrobacter	*A. cumminsii*
Corynebacterium	*C. jeikeium*
Nocardia	*N. brasiliensis*
Bacteroidetes	
Bacteroides	*B. fragilis*
Firmicutes	
Alloiococcus	*A. otitis*
Bacillus	*B. megaterium*
Bacillus	*B. subtilis*
Bacillus	*B. weihenstephanensis*
Enterococcus	*E. avium*
Enterococcus	*E. hirae*
Lactobacillus	*L. jensenii*
Lysinibacillus	*L. fusiformis*
Paenibacillus	*P. durus*
Staphylococcus	*S. arlettae*
Staphylococcus	*S. caprae*
Streptococcus	*S. australis*
Streptococcus	*S. gallolyticus*
Streptococcus	*S. haemolyticus*
Streptococcus	*S. intermedius*
Streptococcus	*S. porcinus*
Proteobacteria	
Acinetobacter	*A. haemolyticus*
Acinetobacter	*A. johnsonii*
Aggregatibacter	*A. segnis*
Alcaligenes	*A. faecalis*
Bordetella	*B. bronchiseptica*
Enterobacter	*E. aerogenes*
Escherichia	*E. hermanii*
Haemophilus	*H. haemolyticus*
Haemophilus	*H. parainfluenzae*
Kluyvera	*K. ascorbata*
Neisseria	*N. subflava*
Neisseria	*N. elongata*
Shewanella	*S. putrifaciens*
Brachyspira	*B. pilosicoli*

studies) to accurately differentiate *S. pneumoniae* from the viridans group of streptococci, which was observed in our laboratory in France, or differentiate some strains of *E. coli* from *Shigella*.[1,6,7,18,21] Thus, conventional biochemical techniques, such as *S. pneumoniae* latex agglutination and indole tests, are sometimes still necessary for accurate identification. Other misidentifications rarely observed but also previously reported were within the genera *Achromobacter*, *Burkholderia*, and *Aeromonas*.[16] Most of these misidentifications can be attributed to an incomplete population of databases associated with the instrument.[1,7,18,21,22] For instance, it has recently been reported that improvements to the database enable a more reliable distinction between *S. pneumoniae* and viridans group streptococci.[23,24] Because the reference databases are not static and expanded regularly to fill in current gaps in identification, updates will continue to improve the performance of MALDI-TOF MS.[18]

To the best of our knowledge, this study represents the first implementation and use of MALDI-TOF MS for the identification of bacteria and fungi in a hospital in tropical Africa. MALDI-TOF MS has previously been implemented and used for research purposes in South Africa for protein identification and bacterial identification, but its use was restricted to a few bacterial species in environmental studies of plant pathology or river water.[25–27] Sample preparation is simple (direct deposit of colonies onto the target slide followed by addition of ready-to-use matrix solution) and can be performed widely. Thus, personnel training requirements are minimal, and samples can be analyzed within minutes. Currently, the estimated wait for one bacterial identification is reported to be from 1 minute and 46 seconds to 2 minutes per sample.[6,16] The use and necessity of this new system were quickly shown by the fact that traditional phenotypic systems were abandoned on the arrival of the MALDI-TOF MS in the laboratory.

TABLE 4
Two hundred six fungi identified at the species level

Phylum, genus, and species	Number of isolates
Ascomycota	
Aspergillus	
A. niger	3
A. flavus	1
Candida	
C. albicans	98
C. tropicalis	42
C. glabrata	30
C. krusei	14
C. parapsilosis	6
C. dubliniensis	2
C. africana	1
C. nivariensis	1
C. utilis	1
Clavispora	
C. lusitaniae	2
Kluyveromyces	
K. marxianus	2
Kodamaea	
K. ohmeri	1
Microsporum	
M. canis	1
Basidiomycota	
Trichosporon	
T. asahii	1

The rapid and accurate identification of routinely encountered bacterial and fungal species as well as those that are rare and difficult to identify using phenotypic methods provides a promising way to improve the care of patients with infectious diseases in Africa. The greatest expenses are associated with purchasing the instrument as well as maintenance fees. The required reagents are not expensive and do not require specific storage conditions if they are prepared in the laboratory.[1,6] Overall, it has been clearly shown that MALDI-TOF MS is less expensive than traditional methods, even when taking into account the costs of reagents, labor, performance measurements, waste disposal, microorganism prevalence, and instrument maintenance expenses as well as multiple runs and additional tests needed to maximize accuracy.[6,18,21,22] Moreover, several studies have shown the clinical benefits of using MALDI-TOF MS. This technique can provide microorganism identification up to 30 hours faster than conventional phenotypic methods; such gains in the speed of identification can have a substantial impact on patient care and management.[6,16,21] MALDI-TOF MS can also be used for the rapid and effective identification of microorganisms from positive blood cultures within 30–45 minutes after a positive signal is provided by a blood culture instrument, and such rapid identification can lead to earlier initiation of treatment with appropriate antimicrobial therapies and increase the chances of obtaining optimal clinical outcomes.[28]

TABLE 5
One hundred nine identifications at the genus level

Phylum and genus	MALDI-TOF MS identification	Number of isolates
Actinobacteria		
Corynebacterium	*Corynebacterium* sp.	1
Firmicutes		
Bacillus	*Bacillus* sp.	2
Enterococcus	*Enterococcus* sp.	2
Lactobacillus	*Lactobacillus* sp.	6
Streptococcus	*S. mitis/oralis/pneumoniae*	29
Streptococcus	*Streptococcus* sp.	19
Proteobacteria		
Achromobacter	*Achromobacter* sp.	10
Acinetobacter	*Acinetobacter* sp.	5
Aeromonas	*Aeromonas* sp.	7
Burkholderia	*Burkholderia* sp.	9
Chryseobacterium	*Chryseobacterium* sp.	1
Citrobacter	*Citrobacter* sp.	10
Ochrobactrum	*Ochrobactrum* sp.	1
Proteus	*P. penneri/vulgaris*	2
Proteus	*Proteus* sp.	1
Pseudomonas	*Pseudomonas* sp.	1
Salmonella	*Salmonella* sp.	3

The primary obstacle to the use of MALDI-TOF MS in Africa is the cost of the machine. In several countries, the consolidation of clinical microbiology laboratories into large core laboratories was intended to lower management costs.[29] The feasibility of MALDI-TOF MS networking in a university hospital in Belgium has recently been shown.[30] We suggest that a common MS platform be developed to be shared among several clinical microbiology laboratories within the city and in nearby areas. The cost of the acquisition can be supported and shared between several organizations, including research organizations, non-governmental organizations, or charity foundations, such as the Mérieux Foundation or the Bill & Melinda Gates Foundation, both of which are already involved in the implementation of new tools to prevent and treat deadly diseases in Africa. Additional costs to be considered relate to equipment maintenance because of the lack of trained personnel in Africa and the cost of spare parts and maintenance contracts.

MALDI-TOF MS also has the potential to identify microorganisms at the subspecies and serotype levels, type strains, and profile antibiotic resistance within minutes.[5,18,31–38] Moreover, MALDI-TOF MS has enabled the rapid detection of tick and mosquito vectors without requiring previous expertise in entomology.[39,40] Thus, this technique will be of use in the implementation of effective prevention measures for vector-borne diseases.[39,40] In the future, it will be useful for tropical countries to evaluate whether MALDI-TOF MS may be used to distinguish between uninfected mosquitoes and those infected with sporozoites of *Plasmodium* spp. and determine potential vector resistance to insecticides.

In the future, an exhaustive repertory of the bacteria correctly identified using MALDI-TOF MS could become available. These data will enable the comparison of bacterial diversity across different areas of the world. They will enable the comparison of rarely observed bacterial species in addition to frequently detected bacteria. For example, comparisons between bacteria observed at the HPD in Dakar and in our laboratory in Marseille show that, during the same time period, among the 10 most frequently identified bacteria at each institution, 8 were found in common: *E. coli*, *K. pneumoniae*, *S. aureus*, *S. epidermidis*, *S. agalactiae*, *E. cloacae*, *P. aeruginosa*, and *E. faecalis* (Figure 1). However, the prevalence rates of these strains were found to be different; *S. agalactiae* was more frequently observed in Dakar, although it has recently become more prevalent in Marseille, whereas *S. epidermidis* was more frequently observed in Marseille.[41] *A. baumannii* and *S. haemolyticus* were among the 10 most frequently identified bacteria only in Dakar, whereas

two *Enterobacteriaceae* (*P. mirabilis* and *E. aerogenes*) were more prevalent in Marseille than Dakar.

MALDI-TOF MS is a single, rapid, robust, and simple-to-use system that has proven its broad applicability and robustness in tropical Africa through its ability to quickly identify a broad range of microorganisms. Despite the initial cost of the MS instrument, the MS technique is more cost-effective than current phenotypic methods, and it would be advantageous to expand the capabilities of the mass MS platform in Africa.

Received June 30, 2014. Accepted for publication December 1, 2014.

Published online January 19, 2015.

Note: Supplemental table appears at www.ajtmh.org.

Acknowledgments: The authors thank Cédric Abat, Thierry Schmidgen, Virginie Giroudon, Dr. Hervé Tissot-Dupont, and the staff of the Hôpital Principal de Dakar (Senegal), especially Dr. Elimane Mbaye and Mr. Diene Bane, for their contributions.

Financial support: This study was supported by the French Ministry of Foreign Affairs, the Research Institute for Development, and the Institut Hospitalo-Universitaire (IHU) Méditerranée Infection.

Authors' addresses: Bécaye Fall, Bissoume Samb-Ba, Silman Diawara, Mamadou Wague Gueye, Kowry Sow, Yaya Diemé, and Boubacar Wade, Hôpital Principal, Dakar, Sénégal, E-mails: becayefall@gmail.com, bissamb@yahoo.fr, diawaraps@hotmail.com, wax304@hotmail.fr, kowryndiaye@yahoo.fr, yaya_dieme@hotmail.com, and bwade55@yahoo.com. Cheikh Ibrahima Lo, Maxence Aubadie-Ladrix, Oleg Mediannikov, Cheikh Sokhna, Didier Raoult, and Florence Fenollar, Aix Marseille Université, Unité des Rickettsies, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63, CNRS 7278, Research Institute for Development 198, Inserm 1095, Marseille, France, and Campus commun UCAD Research Institute for Development, Hann, Dakar, Sénégal, E-mails: cibrahimalo@gmail.com, lamenacexr@yahoo.fr, olegusss1@gmail.com, cheikh.sokhna@ird.fr, didier.raoult@gmail.com, and florence.fenollar@univ-amu.fr. Nadine Perrot and Sonia Chatellier, Département Microbiologie Innovation, bioMérieux, La-Balme-Les-Grottes, France, E-mails: nadine.perrot@biomerieux.com and sonia.chatellier@biomerieux.com.

REFERENCES

1. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D, 2009. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis 49:* 543–551.
2. Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, Francois P, Schrenzel P, 2010. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *J Clin Microbiol 48:* 1169–1175.
3. Stevenson LG, Drake SK, Shea YR, Zelazny AM, Murray PR, 2010. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of clinically important yeast species. *J Clin Microbiol 48:* 3482–3486.
4. Patel R, 2013. Matrix-assisted laser desorption ionization-time of flight mass spectrometry in clinical microbiology. *Clin Infect Dis 57:* 564–572.
5. Seng P, Rolain JM, Fournier PE, La Scola B, Drancourt M, Raoult D, 2010. MALDI-TOF-mass spectrometry applications in clinical microbiology. *Future Microbiol 5:* 1733–1754.
6. Seng P, Abat C, Rolain JM, Colson P, Gouriet F, Fournier PE, Drancourt M, La Scola B, Raoult D, 2013. Emergence of rare pathogenic bacteria in a clinical microbiology laboratory: impact of MALDI-TOF mass spectrometry. *J Clin Microbiol 51:* 2182–2194.
7. Bizzini A, Durussel C, Bille J, Greub G, Prod'hom G, 2010. Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. *J Clin Microbiol 48:* 1549–1554.
8. Chen JH, Yam WC, Ngan AH, Fung AM, Woo WL, Yan MK, Choi GK, Ho PL, Cheng VC, Yuen KY, 2013. Advantages of using MALDI-TOF mass spectrometry as a rapid diagnostic tool for yeast and mycobacteria identification in clinical microbiological laboratory. *J Clin Microbiol 51:* 3981–3987.
9. Itoh E, Watari T, Azuma Y, Watanabe N, Tomoda Y, Akasaka K, Kino S, 2013. Performance of two matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) models for identification of bacteria isolated from blood culture. *Rinsho Byori 61:* 382–389.
10. Panda A, Kurapati S, Samantaray JC, Myneedu VP, Verma A, Srinivasan A, Ahmad H, Behera D, Singh UB, 2013. Rapid identification of clinical mycobacterial isolates by protein profiling using matrix assisted laser desorption ionization-time of flight mass spectrometry. *Indian J Med Microbiol 31:* 117–122.
11. Moon HW, Lee SH, Chung HS, Lee M, Lee K, 2013. Performance of the Vitek MS matrix-assisted laser desorption ionization time-of-flight mass spectrometry system for identification of Gram-positive cocci routinely isolated in clinical microbiology laboratories. *J Med Microbiol 62:* 1301–1306.
12. Nori P, Ostrowsky B, Dorokhova O, Gialanella P, Moy M, Muggia V, Grossberg R, Kornblum J, Lin Y, Levi MH, 2013. Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to resolve complex clinical cases of patients with recurrent bacteremias. *J Clin Microbiol 51:* 1983–1986.
13. Tietz HJ, Hopp M, Schmalreck A, Sterry W, Czaika V, 2001. *Candida africana* sp. nov., a new human pathogen or a variant of *Candida albicans*? *Mycoses 44:* 437–445.
14. Alcoba-Florez J, Mendez-Alvarez S, Cano J, Guarro J, Perez-Roth E, del Pilar AM, 2005. Phenotypic and molecular characterization of *Candida nivariensis* sp. nov., a possible new opportunistic fungus. *J Clin Microbiol 43:* 4107–4111.
15. Eddouzi J, Lohberger A, Vogne C, Manai M, Sanglard D, 2013. Identification and antifungal susceptibility of a large collection of yeast strains isolated in Tunisian hospitals. *Med Mycol 51:* 737–746.
16. Manji R, Bythrow M, Branda JA, Burnham CA, Ferraro MJ, Garner OB, Jennemann R, Lewinski MA, Mochon AB, Procop GW, Richter SS, Rychert JA, Sercia L, Westblade LF, Ginocchio CC, 2014. Multi-center evaluation of the VITEK(R) MS system for mass spectrometric identification of non-enterobacteriaceae gram-negative bacilli. *Eur J Clin Microbiol Infect Dis 33:* 337–346.
17. Alby K, Gilligan PH, Miller MB, 2013. Comparison of MALDI-TOF mass spectrometry platforms for the identification of gram negative rods from cystic fibrosis patients. *J Clin Microbiol 51:* 3852–3854.
18. Clark AE, Kaleta EJ, Arora A, Wolk DM, 2013. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin Microbiol Rev 26:* 547–603.
19. Martiny D, Busson L, Wybo I, El Haj RA, Dediste A, Vandenberg O, 2012. Comparison of the Microflex LT and Vitek MS systems for routine identification of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol 50:* 1313–1325.
20. Garner O, Mochon A, Branda J, Burnham CA, Bythrow M, Ferraro M, Ginocchio C, Jennemann R, Manji R, Procop GW, Richter S, Rychert J, Sercia L, Westblade L, Lewinski M, 2014. Multi-centre evaluation of mass spectrometric identification of anaerobic bacteria using the VITEK MS system. *Clin Microbiol Infect 20:* 335–339.
21. Tan KE, Ellis BC, Lee R, Stamper PD, Zhang SX, Carroll KC, 2012. Prospective evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: a bench-by-bench study for assessing the impact on time to identification and cost-effectiveness. *J Clin Microbiol 50:* 3301–3308.
22. Gaillot O, Blondiaux N, Loiez C, Wallet F, Lemaitre N, Herwegh S, Courcol RJ, 2011. Cost-effectiveness of switch to matrix-assisted laser desorption ionization-time of flight mass

spectrometry for routine bacterial identification. *J Clin Microbiol 49:* 4412.
23. Werno AM, Christner M, Anderson TP, Murdoch DR, 2012. Differentiation of *Streptococcus pneumoniae* from nonpneumococcal streptococci of the *Streptococcus mitis* group by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol 50:* 2863–2867.
24. Dubois D, Segonds C, Prere MF, Marty N, Oswald E, 2013. Identification of clinical *Streptococcus pneumoniae* isolates among other alpha and nonhemolytic streptococci by use of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system. *J Clin Microbiol 51:* 1861–1867.
25. Standing TA, du Plessis E, Duvenage S, Korsten L, 2013. Internalisation potential of *Escherichia coli* O157:H7, *Listeria monocytogenes*, *Salmonella enterica* subsp. *enterica* serovar *Typhimurium* and *Staphylococcus aureus* in lettuce seedlings and mature plants. *J Water Health 11:* 210–223.
26. Gemmell ME, Schmidt S, 2013. Is the microbiological quality of the Msunduzi River (KwaZulu-Natal, South Africa) suitable for domestic, recreational, and agricultural purposes? *Environ Sci Pollut Res Int 20:* 6551–6562.
27. Haarburger D, Bergstrom J, Pillay TS, 2013. Serum proteome changes following human immunodeficiency virus infection. *Clin Lab 59:* 639–646.
28. Foster AG, 2013. Rapid identification of microbes in positive blood cultures by matrix-assisted laser desorption/ionisation time-of-flight (Maldi-Tof) mass spectrometry- (Vitek MS–bioMerieux.). *J Clin Microbiol 51:* 3717–3719.
29. Fournier PE, Drancourt M, Colson P, Rolain JM, La Scola B, Raoult D, 2013. Modern clinical microbiology: new challenges and solutions. *Nat Rev Microbiol 11:* 574–585.
30. Martiny D, Cremagnani P, Gaillard A, Miendje Deyi VY, Mascart G, Ebraert A, Attalibi S, Dediste A, Vandenberg O, 2014. Feasibility of matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) networking in university hospitals in Brussels. *Eur J Clin Microbiol Infect Dis 33:* 745–754.
31. Hrabak J, Walkova R, Studentova V, Chudackova E, Bergerova T, 2011. Carbapenemase activity detection by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol 49:* 3222–3227.
32. Walker J, Fox AJ, Edwards-Jones V, Gordon DB, 2002. Intact cell mass spectrometry (ICMS) used to type methicillin-resistant *Staphylococcus aureus*: media effects and inter-laboratory reproducibility. *J Microbiol Methods 48:* 117–126.
33. Jackson KA, Edwards-Jones V, Sutton CW, Fox AJ, 2005. Optimisation of intact cell MALDI method for fingerprinting of methicillin-resistant *Staphylococcus aureus*. *J Microbiol Methods 62:* 273–284.
34. Du Z, Yang R, Guo Z, Song Y, Wang J, 2002. Identification of *Staphylococcus aureus* and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Chem 74:* 5487–5491.
35. Majcherczyk PA, McKenna T, Moreillon P, Vaudaux P, 2006. The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus aureus*. *FEMS Microbiol Lett 255:* 233–239.
36. Camara JE, Hays FA, 2007. Discrimination between wild-type and ampicillin-resistant *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Bioanal Chem 389:* 1633–1638.
37. Sparbier K, Schubert S, Weller U, Boogen C, Kostrzewa M, 2012. Matrix-assisted laser desorption ionization-time of flight mass spectrometry-based functional assay for rapid detection of resistance against beta-lactam antibiotics. *J Clin Microbiol 50:* 927–937.
38. Kostrzewa M, Sparbier K, Maier T, Schubert S, 2013. MALDI-TOF MS: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. *Proteomics Clin Appl 7:* 767–778.
39. Yssouf A, Socolovschi C, Flaudrops C, Ndiath MO, Sougoufara S, Dehecq JS, Lacour G, Berenger JM, Sokhna CS, Raoult D, Parola P, 2013. Matrix-assisted laser desorption ionization—time of flight mass spectrometry: an emerging tool for the rapid identification of mosquito vectors. *PLoS ONE 8:* e72380.
40. Yssouf A, Flaudrops C, Drali R, Kernif T, Socolovschi C, Berenger JM, Raoult D, Parola P, 2013. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of tick vectors. *J Clin Microbiol 51:* 522–528.
41. Abat C, Chaudet H, Raoult D, Colson P, 2014. Increasing trend of invasive group B streptococcal infections, Marseille, France. *Clin Infect Dis 58:* 750–751.

SUPPLEMENTAL TABLE 1

List of the 93 human bacterial isolates tested to validate the use of VITEK® MS RUO for diagnostic purposes

Bacteria	Number of isolates
Acinetobacter baumannii	7
Citrobacter koseri	2
Enterobacter aerogenes	1
Enterobacter cloacae	2
Escherichia coli	13
Enterococcus faecium	2
Enterococcus faecalis	8
Klebsiella oxytoca	3
Klebsiella pneumoniae	10
Morganella morganii	1
Proteus mirabilis	2
Pseudomonas aeruginosa	2
Streptococcus agalactiae	8
Staphylococcus aureus	14
Staphylococcus epidermidis	10
Stenotrophomonas maltophilia	5
Streptococcus pyogenes	3

Conclusion et Perspective tirées du Chapitre 2:

Les travaux effectués pendant cette étude montrent que le MALDI-TOF est un système simple, rapide et facile, qui a prouvé son applicabilité et sa robustesse en Afrique tropicale dans l'identification des microorganismes à large échelle. En outre, le délai d'identification complète d'une bactérie par MALDI-TOF est estimé à une minute quarante six secondes à deux minutes maximum. Il faut noter aussi que l'utilisation de l'appareil ne demande pas un grand effectif ni une grande expertise.

L'installation du MALDI-TOF à l'hôpital Principal de Dakar en 2012 est une première pour un laboratoire de microbiologie clinique situé en Afrique de l'Ouest. En effet ce retard était dû au coût d'investissement élevé de l'appareil. Ainsi avec la possibilité d'obtenir des financements auprès d'organismes partenaires, un tel appareil peut être acquis sans grande difficulté. Reste maintenant à relever le défi de certaines contraintes secondaires comme la maintenance et la bonne utilisation de l'instrument. L'identification rapide, précise et à haut débit des bactéries et champignons rencontrés dans ce laboratoire clinique, tels que les microorganismes rares et difficiles à identifier par les conventionnelles méthodes phénotypiques, montre que le MALDI-TOF est une nouvelle alternative pour améliorer la prise en charge des patients en réduisant ainsi le délai de rendu des résultats de prélèvements et par conséquent facilitant la lutte contre les maladies infectieuses émergentes en Afrique.

A l'avenir un répertoire complet des bactéries correctement identifiées par MALDI-TOF en Afrique sera disponible grâce à une augmentation du nombre d'appareils dans cette zone tropicale. En effet, ces données recueillies permettront la comparaison de la diversité bactérienne entre différentes régions du monde mais aussi et surtout une comparaison des espèces de bactéries rarement observées à celles fréquemment rencontrées, ce qui permettra de mieux comprendre l'épidémiologie de certains agents pathogènes. Par exemple, la comparaison des souches identifiées pendant la même période (10 mois) entre l'Hôpital Principal de Dakar et le laboratoire de l'Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE) de Marseille a permis de connaitre les dix bactéries les plus fréquemment isolées, dont huit sont communes aux deux laboratoires à savoir : *Escherchia coli, Klebsiella pneumoniae, Staphylococcus aureus, Staphylococcus epidermidis, Streptococcus agalactiae, Enterobacter cloacae, Pseudomonas aeruginosa,* et *Enterobacter faecalis.* Il est aussi important en Afrique de penser au partage de la plateforme MALDI-TOF par plusieurs structures publiques ou privées de santé situées à des distances raisonnables (< 20 km).

Chapitre 3 :

Description des nouvelles espèces de bactéries isolées au cours de notre travail de thèse

Généralités sur la taxonomie bactérienne

La taxonomie ou la systématique est la science qui permet de classer par groupe des individus partageant un certain nombre de critères morphologiques. L'unité fondamentale de la taxonomie en microbiologie médicale est **l'espèce**. «L'espèce bactérienne est par définition une population de cellules ayant des caractéristiques similaires, correspondant à une souche précise dite souche type» (35). Comme les bactéries se divisent en permanence, la notion de **souche** est introduite pour désigner une population de bactéries issue d'une même cellule (36). Actuellement les microbiologistes s'appuient sur plusieurs techniques d'identification pour déterminer l'appartenance d'une souche bactérienne à une espèce donnée. En effet, les espèces sont souvent définies selon une approche numérique (37), une approche phylogénétique ou une approche polyphasique. La taxonomie numérique consiste à utiliser plusieurs critères biochimiques, morphologiques et culturaux ainsi que les sensibilités aux antibiotiques et des composants inorganiques dans le but de déterminer le degré de similarité entre les individus (38). L'approche phylogénétique est une méthode qui permet de mesurer le nombre de séquences d'ADN qu'ont en commun deux

individus et d'en déduire le pourcentage de divergence entre ces séquences qui sont proches mais non identiques (38). La taille du génome, le contenu guanine-plus-cytosine (G + C), et la stabilité thermique de séquences d'ADN apparentées, sont parmi les facteurs les plus souvent étudiés lors de l'approche phylogénétique(39). Enfin dans la pratique l'approche de la taxonomie bactérienne doit être polyphasique car elle permet d'identifier et de classer les bactéries sur tous les niveaux en suivant principalement trois étapes:

- groupage phénotypique des souches selon des caractéristiques biochimiques, morphologiques et autres;
- déterminer dans ce groupe si l'homogénéité (ou hétérogénéité) phénotypique observée correspond à l'homogénéité (ou hétérogénéité) phylogénétique après des tests biomoléculaires (approche phylogénétique)
- enfin l'étape la plus importante est un réexamen des caractéristiques biochimiques des groupes à ADN proche.

Dans notre travail, une souche est suspectée être une nouvelle espèce lorsqu'elle apparait inconnue des bases de données du MALDI-TOF et que son pourcentage de similarité avec les séquences de l'ARN 16S ribosomal est inférieur à 98,7% (40). En effet, notre première intention lorsqu'une bactérie n'est pas identifiée pendant plusieurs tentatives (deux à trois fois) par la spectrométrie de masse MALDI-TOF, est de procéder à l'amplification et au séquençage du gène de l'ARN 16S ribosomique (le plus souvent) ou du gène de ménage *rpoB* (quand on connait la famille de la bactérie à l'avance).

En réalité les bactéries possèdent des ribosomes composés d'une grande sous-unité (50S) et d'une petite sous-unité (30S) (41). L'ARN ribosomal (ARNr) 16S est la composante de la petite sous unité ribosomale des procaryotes. Il est codé par le gène qui porte son nom «16S rRNA gene » (42). Ce gène est essentiellement utilisé par les microbiologistes en raison de sa structure, très conservée dans toutes les bactéries et du nombre de copies variables qu'il peut exister (42,43). En outre ce gène de l'ARNr 16S a une longueur d'environ 1500 nucléotides et présente en même temps sept régions conservées et neuf régions hypervariables (43). C'est pour ces raisons qu'il a été recommandé de séquencer au moins 500 à 525 bp, idéalement 1300 à 1500 bp dans le processus d'identification des bactéries (44,45). Les séquences du gène 16S ont largement participé à la découverte de nouvelles espèces de bactéries ces dernières décennies en se basant sur l'étude de la phylogénie bactérienne (46,47). D'ailleurs des amorces spécifiques appelées «amorces universelles» ont été conçues pour pouvoir amplifier toute la longueur du gène qui code pour l'ARNr 16S de la plupart des bactéries (48).

Malgré l'universalité du gène et son état conservé, son utilisation pour la détermination d'une nouvelle espèce est très controversée du fait de l'absence de consensus sur une valeur seuil (46,49). En revanche, le comité Stackebrandt stipule que toute description de nouvelle espèce devait comporter une séquence complète de l'ARNr 16S par opposition à la traditionnelle méthode de l'hybridation ADN-ADN, qui est une technique lourde et complexe (50). Par ailleurs, si on se réfère à la cinétique de réassociation ADN-ADN, une espèce est ainsi définie

génétiquement comme le regroupement de souches qui partagent des liens ADN-ADN selon un pourcentage d'hybridation ADN-ADN ≥ 70% et une stabilité thermique des hybrides ≤ 5°C (46).

Toutes les nouvelles espèces décrites au cours de nos travaux suivent une approche polyphasique qui repose sur des analyses de génome, de spectres de MALDI-TOF ainsi que des critères biochimiques de la bactérie (51).

Article 4:

« Non-contiguous finished genome sequence and description of Clostridium dakarense sp. nov. »

Cheikh Ibrahima Lo[1], Ajay Kumar Mishra[1], Roshan Padhmanabhan[1], Bissoume Samb Ba[2], Amy Gassama Sow[2], Catherine Robert[1], Carine Couderc[1], Ngor Faye[3], Didier Raoult[1], Pierre-Edouard Fournier[1], and Florence Fenollar[1]*

[1]Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), Marseille, France

[2]Unité de Bactériologie Expérimentale Institut Pasteur de Dakar 36, Dakar, Senegal

[3]Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann, Senegal

Publié dans:

Standards in Genomic Sciences.

September 2013; Volume 9; Issue 1; pages: 14-27.

http://www.ncbi.nlm.nih.gov/pubmed/24501642

Résumé de l'article 4: ***« Non-contiguous finished genome sequence and description of Clostridium dakarense sp. nov. »***

La souche *Clostridium dakarense* FF1^T, est la souche type de *Clostridium dakarense* sp. nov., une nouvelle espèce du genre *Clostridium*. Cette souche a été isolée de la flore fécale d'une enfant Sénégalaise de 4 mois souffrant de gastroentérite à Dakar au Sénégal. Les séquences de l'ARN 16S ribosomique et du génome de *C. dakarense* souche FF1^T (= CSUR P243 = DSM 27086) ont été déposées dans GenBank avec les numéros d'accession respectifs KC517358 et CBTZ00000000. La souche type de *C. dakarense* est FF1^T.

La souche *C. dakarense* sp. nov. FF1^T est un bacille Gram positif anaérobie strict. Aucune croissance n'est observée en condition aérobie. Les bactéries sont en forme de bâtonnet avec un diamètre moyen de 1,2 µm. La bactérie forme des spores et elle est mobile. La croissance de la bactérie peut être observée entre 25 et 37°C, mais la croissance optimale est observée à 37°C sur gélose Columbia enrichi avec 5% de sang de mouton. Les colonies ont un diamètre de 1,5 mm sur gélose Columbia enrichi au sang et sur gélose Chocolat PolyViteX.

Les activités de catalase, oxydase, uréase, indole et nitrate réduction sont absentes. Les activités d'arginine dihydrolase, N-acetyl-β-glucosanimidase et acide pyroglutamique arylamidase sont présentes. La bactérie est sensible *in vitro* à l'amoxicilline, au métronidazole, à la vancomycine, à l'imipénème et à la rifampicine mais résistante au triméthoprime/sulfaméthoxazole.

Le génome de 3 735 762 paires de bases composé d'un chromosome (mais pas de plasmide), a un contenu en G+C de 27,98% et contient 3 843 protéines codantes et 73 gènes ARN, y compris 8 gènes ARNr.

Description de *Clostridium dakarense* sp. nov

Sur la base des analyses phénotypique, phylogénétique et génomique, nous proposons la création de *Clostridium dakarense* sp. nov. qui contient la souche type $FF1^T$. Cette bactérie a été isolée de la flore fécale d'une enfant Sénégalaise de 4 mois souffrant de gastroentérite.

Clostridium dakarense (da.kar.e′n.se. L. gen. neutr. n. *dakarense*, de Dakar, la capitale du Sénégal, où la souche type a été isolée).

Non-contiguous finished genome sequence and description of *Clostridium* dakarense sp. nov.

Cheikh Ibrahima Lo[1], Ajay Kumar Mishra[1], Roshan Padhmanabhan[1],Bissoume Samb Ba[2], Amy Gassama Sow[2], Catherine Robert[1], Carine Couderc[1], Ngor Faye[3], Didier Raoult[1], Pierre-Edouard Fournier[1], and Florence Fenollar[1*]

[1]Aix-Marseille Université, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), Marseille, France

[2]Unité de Bactériologie Expérimentale Institut Pasteur de Dakar 36, Dakar, Senegal

[3]Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann (Senegal)

***Correspondence**: Florence Fenollar (florence.fenollar@univ-amu.fr)

Keywords: *Clostridium dakarense*, genome, culturomics, taxono-genomics

Clostridium dakarense strain FF1^T, is the type strain of *Clostridium dakarense* sp. nov., a new species within the genus *Clostridium*. This strain, whose genome is described here, was isolated from the fecal flora of a 4-month-old Senegalese child suffering from gastroenteritis. *C. dakarense* sp. nov. strain FF1^T is an obligate anaerobic Gram-positive bacillus. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 3,735,762 bp long genome (1 chromosome but no plasmid) exhibits a G+C content of 27.98% and contains 3,843 protein-coding and 73 RNA genes, including 8 rRNA genes.

Introduction

Clostridium dakarense strain FF1^T (= CSUR P243 = DSM 27086), is the type strain of *Clostridium dakarense* sp. nov. This bacterium is a Gram-positive, anaerobic, spore-forming, indole negative bacillus that was isolated from the stool of a 4-month-old Senegalese child suffering from gastroenteritis as part of a "culturomics" study aiming at cultivating individually all species within human feces.

The elevated cost and lack of intra- and inter-laboratory reproducibility of the "gold standard" of taxonomic tools. i.e. DNA-DNA hybridization and G+C content determination [1], put bacterial taxonomic classification in a precarious state. In addition, the internationally-validated cutoff values of 16S rRNA sequence comparison [2] do not apply to all validly published genera and species. Recently, high throughput genome sequencing and mass spectrometric analyses of bacteria have allowed unprecedented access to a wealth of genetic and proteomic information [3]. As a consequence, we proposed to use a polyphasic approach [4] to describe new bacterial taxa, including genome sequence, MALDI-TOF spectrum and main phenotypic characteristics [5-11].

The genus *Clostridium* (Prazmowski, 1880), classified among the *Firmicutes*, was created in 1880 [12] and consists of obligate anaerobic rod-shaped bacilli capable of producing endospores [12]. More than 180 *Clostridium* species have been described to date [13]. Members of the genus *Clostridium* are mostly environmental bacteria or associated with the commensal digestive flora of mammals, but several are major human pathogens, including *C. botulinum*, *C. difficile*, *C. tetani* and *C. perfringens* [14,15]. A few species, such as *C. butyricum* and *C. pasteurianum*, fix nitrogen and have gained importance in agricultural and industrial applications [16,17].

Here we present a summary classification and a set of features for *C. dakarense* sp. nov. strain FF1^T (= CSUR P243 = DSM 27086) together with the description of the complete genomic sequencing and annotation. These characteristics support the circumscription of the species *C. dakarense* sp. nov.

Classification and features

A stool specimen was collected from a 4-month-old Senegalese child suffering from gastroenteritis. Informed consent was obtained from the child's parents and approval from the ethics committee from the Institut Federatif de Recherche 48 (Faculté de Médecine, Marseille, France). The fecal specimen was preserved at -20°C after collection and sent to Marseille. Strain FF1^T (Table 1) was isolated in July 2011 by anaerobic cultivation on 5% sheep blood-enriched Columbia agar (BioMerieux, Marcy l'Etoile, France). This strain exhibited a 96.90% 16S rRNA nucleotide sequence similarity with *C. lituseburense*, the phylogenetically closest validated *Clostridium* species (Figure 1). Although sequence similarity of the 16S rRNA is not uniform across taxa, this value was lower than the 98.7% threshold recommended by Stackebrandt and Ebers to delineate a new species without carrying out DNA-DNA hybridization [30]. In addition, it was consistent with 16S rRNA identity values observed among validated species within the *Clostridium* genus that range from 78.4 to 99.8%.

Table 1. Classification and general features of *Clostridium dakarense* strain FF1^T according to the MIGS recommendations [18].

MIGS ID	Property	Term	Evidence code[a]
		Domain *Bacteria*	TAS [19]
		Phylum *Firmicutes*	TAS [20-22]
		Class *Clostridia*	TAS [23,24]
	Current classification	Order *Clostridiales*	TAS [25,26]
		Family *Clostridiaceae*	TAS [25,27]
		Genus *Clostridium*	TAS [12,25,28]
		Species *Clostridium dakarense*	IDA
		Type strain FF1	IDA
	Gram stain	Positive	IDA
	Cell shape	Rod-shaped	IDA
	Motility	Motile	IDA
	Sporulation	Sporulating	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37°C	IDA
MIGS-6.3	Salinity	Growth in BHI medium + 5% NaCl	IDA
MIGS-22	Oxygen requirement	Anaerobic	IDA
	Carbon source	Unknown	NAS
	Energy source	Unknown	NAS
MIGS-6	Habitat	Human gut	IDA
MIGS-15	Biotic relationship	Free living	IDA
	Pathogenicity	Unknown	
	Biosafety level	2	
MIGS-14	Isolation	Human feces	NAS
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection time	June 2011	IDA
MIGS-4.1	Latitude	13.7167	IDA
MIGS-4.1	Longitude	- 16.4167	IDA
MIGS-4.3	Depth	Surface	IDA
MIGS-4.4	Altitude	51 m above sea level	IDA

[a]Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [29]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

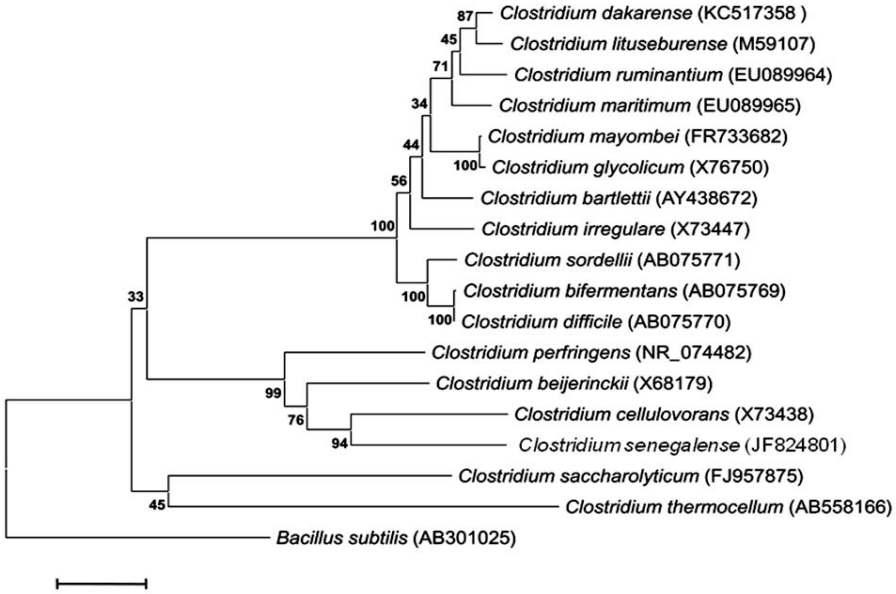


Figure 1. Phylogenetic tree highlighting the position of *C. dakarense* sp. nov. strain FF1^T relative to other type strains within the *Clostridium* genus. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences obtained using the maximum-likelihood method within the MEGA software. Numbers at the nodes are bootstrap values obtained by repeating 500 times the analysis to generate a majority consensus tree. *Bacillus subtilis* was used as an outgroup. The scale bar represents a 2% nucleotide sequence divergence.

Different growth temperatures (25, 30, 37, 45 and 56°C) were tested. Growth was observed between 25 and 37°C, with optimal growth at 37°C after 24 hours of inoculation in anaerobic conditions. Colonies were 1.5 mm in diameter and opaque and smooth appearance on blood-enriched Columbia agar. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMerieux), and under aerobic conditions, with or without 5% CO_2. The strain growth was obtained only in anaerobic conditions. Gram staining showed rod-shaped Gram-positive bacilli able to form spores (Figure 2). The motility test was positive. Cells grown on agar have a mean diameter of 1.2 µm (Figure 3).

Strain FF1^T exhibited neither catalase nor oxidase activities. Using API Rapid ID 32A (BioMerieux, Marcy l'Etoile), a positive reaction were observed for arginine dihydrolase, N-acetyl-β-glucosaminidase and pyroglutamic acid arylamidase. Negative reactions were observed for urease, indole and nitrate reduction. Using API 50 CH (BioMerieux, Marcy l'Etoile), positive reactions were observed for galactose, glucose, maltose and saccharose fermentation and negative reaction were observed for ribose, lactose and fructose. *C. dakarense* is susceptible to amoxicillin, metronidazole, vancomycin, imipenem and rifampicin and resistant to trimethoprim/ sulfamethoxazole. When compared with representative species from the genus *Clostridium*, *C. dakarense* strain FF1^T exhibited the phenotypic differences detailed in Table 2.

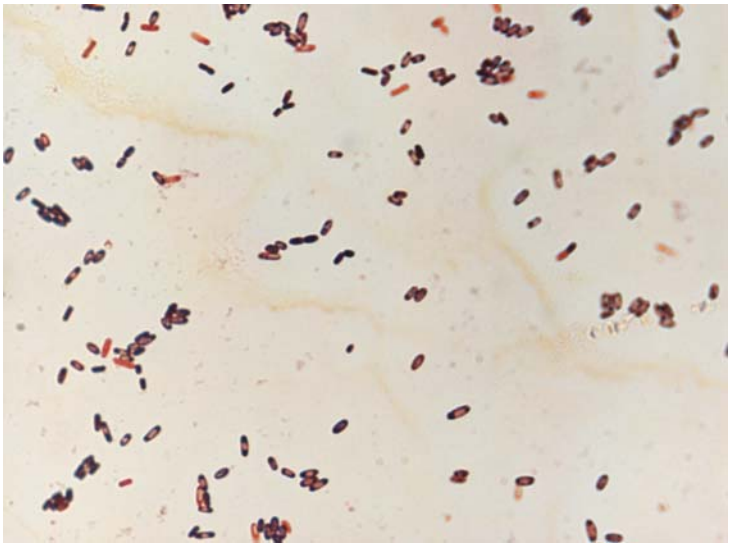


Figure 2. Gram staining of *C. dakarense* sp. nov. strain FF1^T.

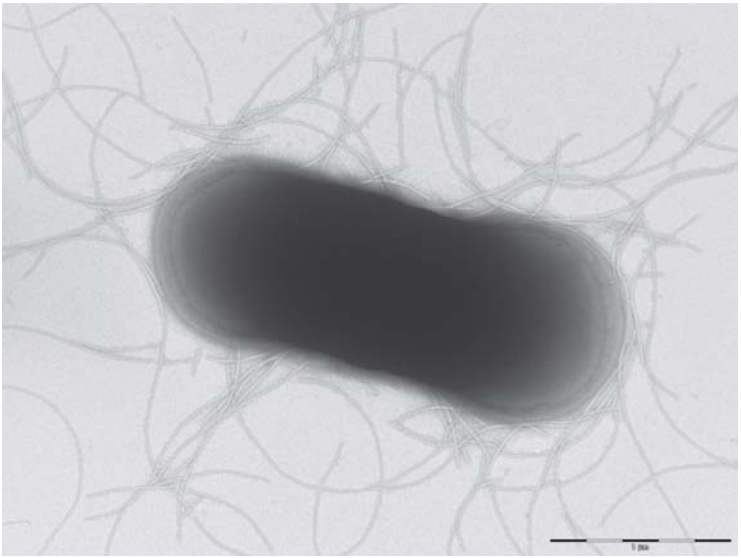


Figure 3. Transmission electron microscopy of *C. dakarense* sp. nov. strain FF1^T, using a Morgani 268D (Philips) at an operating voltage of 60kV. The scale bar represents 1 µm.

Table 2. Differential characteristics of *C. dakarense* sp. nov. strain FF1^T (Cda)

Properties	**CDa**	**CBa**	**CBe**	**CC**	**CDi**	**CG**	**CP**	**CSa**	**CSe**	**CT**
Cell diameter (µm)	1.2	1.5	1.7	2.5	3.0	0.4-1.0	1.3	3.0	1.1	2.5
Oxygen requirement	-	-	-	-	-	-	-	-	-	na
Pigment production	-	-	-	-	+	+	+	na	-	+
Gram stain	+	+	V	-	+		+	-	+	-
Salt requirement	-	na	na	na	na	-	-	na	-	na
Motility	+	-	+	-	+	+	-	-	+	-
Endospore formation	+	+	+	+	+	+	w	+	+	+
Production of										
Acid phosphatase	+	+	na	na	na	na	+	na	na	na
Catalase	-	-	-	-	na	na	na	na	-	na
Oxidase	-	na	na	na	na	na	na	na	-	na
Nitrate reductase	-	-	-	na	-	-	+	+	-	-
Urease	-	-	-	na	na	na	na	na	-	na
β-galactosidase	-	+	na	na	na	-	+	na	-	na
Acid from										
L-Arabinose	-	na	+	-	-	-	-	+	na	na
Ribose	-	+	-		-	-	+	w	na	na
Mannose	-	-	+		+	-	+	na	na	na
Mannitol	-	+	+	+	+	-	-	w	na	na
Sucrose	-	+	+	+	+	-	+	w	na	na
D-glucose	+	+	+	+	na	+	+		na	na
D-fructose	-	+	+	+	+	+	+	+	na	na
D-maltose	+	+	+	+	-	+	+	w	na	na
D-lactose	-	na	+	+	-	-	+	w	na	na
Hydrolysis of										
Gelatin	na	-	+	-	na	-	na	na	na	+
Starch	na	na	+	-	-	-	+	-	na	
G+C content (mol%)	27.98	29.8	28	27	28	29	27	28	26.8	39
Habitat	Human gut	Human gut	Human gut	Poplar wood	Human gut	Mud, wastewater	Colonic flora	Sewage sludge	Human gut	Sewage sludge

C. bartlettii (CBa), *C. beijerinckii* (CBe), *C. cellulovorans* (CC), *C. difficile* (CDi), *C. glycolicum* (CG), *C. perfringens* (CP), *C. saccharolyticum* (CSa), *C. senegalense* (CSe) and *C. thermocellum* (CT).

na = data not available; w = weak, v = variable reaction

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [31]. Briefly, a pipette tip was used to pick one isolated bacterial colony from a culture agar plate, and to spread it as a thin film on a MTP 384 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany). Eighteen distinct deposits were made for strain FF^T from eighteen isolated colonies. Each smear was overlaid with 2 µL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile, 2.5% tri-fluoracetic-acid, and allowed to dry for five minutes. Measurements were performed with a Microflex spectrometer (Bruker). Spectra were recorded in the positive linear mode for the mass range of 2,000 to 20,000 Da (parameter settings: ion source 1 (IS1), 20 kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 675 shots at a variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. The eighteen spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against the main spectra of 4,706 bacteria including 216 spectra from validly published species of *Clostridium*, that are part of the reference data contained in the BioTyper database. The method of identification included the m/z from 2,000 to 20,000 Da. For every spectrum, 100 peaks at most were taken into account and compared with spectra in the database. A score enabled the identification, or not, from the tested species: a score > 2 with a validly published species enabled the identification at the species level, and a score < 1.7 did not enable any identification at the genus level. For strain FF1^T, the maximal obtained score was lower than 1.9, thus suggesting that our isolate was not a member of a known species. We added the spectrum from strain FF1^T to our database for future reference (Figure 4). Finally, the gel view allows us to highlight the spectrum differences with other members of the genus *Clostridium* (Figure 5).

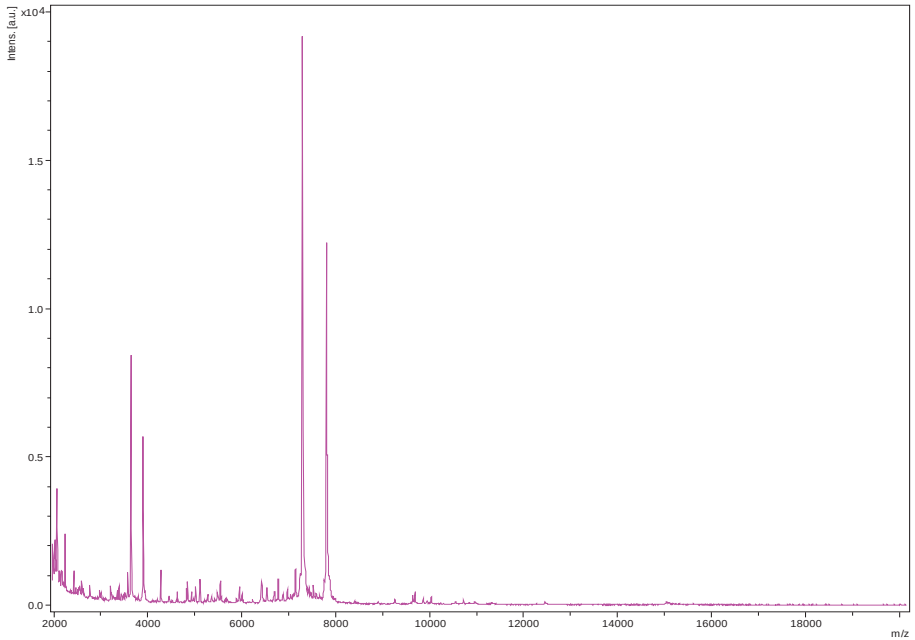


Figure 4. Reference mass spectrum from *C. dakarense* strain FF1^T. Spectra from 18 individual colonies were compared and a reference spectrum was generated.

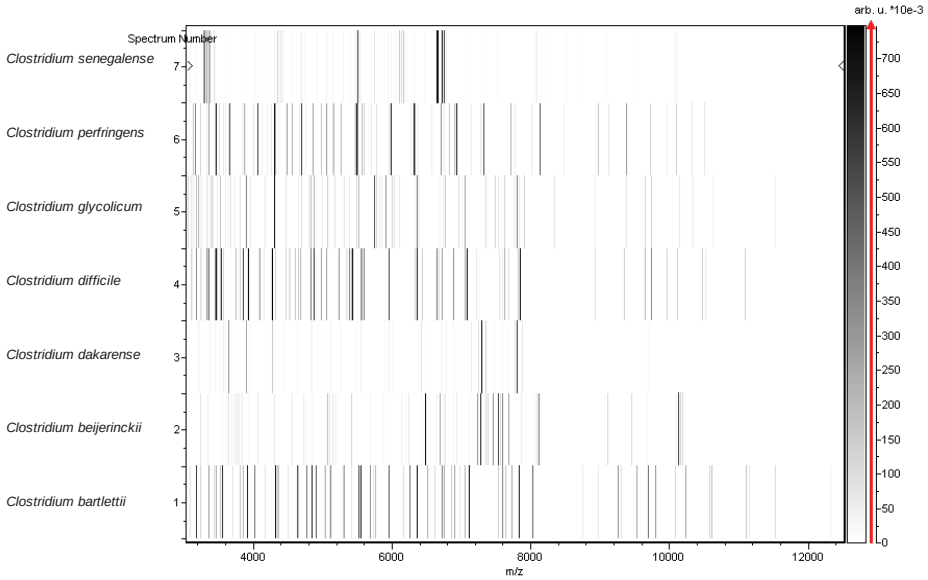


Figure 5. Gel view comparing *C. dakarense* sp. nov. strain FF1^T spectra with other members of the *Clostridium* genus (*C. bartlettii*, *C. beijerinckii*, *C. difficile*, *C. glycolicum*, *C. perfringens*, *C. senegalense*). The Gel View displays the raw spectra of all loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Gray scale scheme code. The color bar and the right y-axis indicate the relation between the color a peak is displayed with and the peak intensity in arbitrary units.

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its phylogenetic position and 16S rRNA similarity to other members of the genus *Clostridium*, and is part of a "culturomics" study of the human digestive flora aiming at isolating all bacterial species within human feces. It was the 94th genome of a *Clostridium* species and the first genome of *Clostridium dakarense* sp. nov. The Genbank accession number is CBTZ00000000 and consists of 257 contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [32].

Growth conditions and DNA isolation

C. dakarense sp. nov. strain FF1^T (= CSUR P243 = DSM 27086), was grown anaerobically on sheep blood-enriched Columbia agar medium at 37°C. Eight petri dishes were spread and resuspended in 4x100µl of G2 buffer (EZ1 DNA Tissue kit, Qiagen). A first mechanical lysis was performed by glass powder on the Fastprep-24 device (Sample Preparation system) from MP Biomedicals, USA) using 2x20 seconds cycles. DNA was then treated with 2.5 µg/µL lysozyme (30 minutes at 37°C) and extracted through the BioRobot EZ 1 Advanced XL (Qiagen). The DNA was then concentrated and purified on a Qiamp kit (Qiagen). DNA concentration was 70.7ng/µl as determined by the Genios Tecan fluorometer, using the Quant-it Picogreen kit (Invitrogen).

Table 3. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	One 454 paired end 3-kb library
MIGS-29	Sequencing platforms	454 GS FLX Titanium
MIGS-31.2	Fold coverage	35
MIGS-30	Assemblers	Newbler version 2.5.3
MIGS-32	Gene calling method	Prodigal
	Genbank ID	CBTZ00000000
	Genbank Date of Release	
MIGS-13	Project relevance	Study of the human gut microbiome

Genome sequencing and assembly

This project was loaded twice on a 1/4 region for the paired-end application and once on a 1/8 region for the shotgun on PTP Picotiterplates. The shotgun library was constructed with 500 ng of DNA as described by the manufacturer (Roche) with the GS Rapid library Prep kit. For the paired-end sequencing, 5 µg of DNA was mechanically fragmented on the Hydroshear device (Digilab, Holliston, MA, USA) with an enrichment size of 3-4kb. The DNA fragmentation was visualized using an Agilent 2100 BioAnalyzer on a DNA labchip 7500, which yield an optimal size of 3.6 kb. The library was constructed according to the 454_Titanium paired-end protocol and manufacturer. Circularization and nebulization were performed and generated a pattern with an optimum at 561 bp. After PCR amplification through 15 cycles followed by double size selection, the single stranded paired end library was then quantified with Quant-it Ribogreen kit (Invitrogen) on the Genios_Tecan fluorometer at 52 pg/µL. The library concentration equivalence was calculated as 1.7E+08 molecules/µL. The library was stored at -20°C until use.

The shotgun library was clonally amplified with 3cpb in 3 emPCR reactions and the paired end library was amplified with lower cpb (1cpb) in 4 emPCR reactions with the GS Titanium SV emPCR Kit (Lib-L) v2. The yield of the emPCR was 5.37% for the shotgun reads and 19.27% for the paired-end reads, according to the quality expected by the range of 5 to 20% from the Roche procedure. A total of 340,000 beads from the 1/8 region of the shotgun reads and 790,000 beads from the 1/4 region of the paired-end reads were loaded on the GS Titanium PicoTiterPlates (PTP Kit 70×75) and sequenced with the GS Titanium Sequencing Kit XLR70.

The runs were performed overnight and then analyzed on the cluster through the gsRunBrowser and gsAssembler_Roche. The global 383,079 passed filter sequences generated 96.50 Mb with a length average of 277 bp. These sequences were assembled using the Newbler software from Roche with 90% identity and 40 bp as overlap. Fourteen scaffolds and 257 large contigs (>1500bp) were obtained, for a genome size of 3,735,762 bp.

Genome annotation

Open Reading Frames (ORFs) were predicted using Prodigal [33] with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank database [34] and the Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAScanSE tool [35] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [36] and BLASTn against the GenBank database. Lipoprotein signal peptides and numbers of transmembrane helices were predicted using SignalP [37] and TMHMM [38] respectively. ORFans were identified if their BLASTP *E*-value was lower than $1e^{-03}$ for alignment length greater than 80 amino acids. If alignment lengths

were smaller than 80 amino acids, we used an *E*-value of 1e-05. Such parameter thresholds have already been used in previous works to define ORFans. Artemis [39] was used for data management and DNA Plotter [40] was used for visualization of genomic features. Mauve alignment tool was used for multiple genomic sequence alignment and visualization [41].

To estimate the mean level of nucleotide sequence similarity at the genome level between *C. dakarense* and nine other members of the genus *Clostridium* (Table 6), we use the Average Genomic Identity of gene Sequences (AGIOS) homemade software. Briefly, this software combines the Proteinortho software [42] for detecting orthologous proteins between genomes compared two by two, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm. *Clostridium* ***dakarense*** strain FF1^T, was compared to *C. bartlettii* strain DSM 16795 (GenBank accession number NZ_DS499569), *C. beijerinckii* strain NCIMB 8052 (NC_009617), *C. cellulovorans* strain 743B (NC_014393), *C. difficile* strain 630 (NC8009089), *C. glycolicum* strain ATCC 14880 (ARES01000000), *C. perfringens* strain ATCC 13124 (BA000016), *C. saccharolyticum* strain WM1 (NC_014376), *C. senegalense* strain JC122^T (CAEV00000000), and *C. thermocellum* strain ATCC 27405 (CP000568).

Genome properties

The genome of *C. dakarense* sp. nov. strain FF1^T is 3,735,762 bp long (1 chromosome, but no plasmid) with a 27,98% G + C content of (Figure 6 and Table 4). Of the 3,916 predicted genes, 3,843 protein-coding genes, and 73 were RNAs. Eight rRNA genes (one 16S rRNA, one 23S rRNA and six 5S rRNA) and 65 predicted tRNA genes were identified in the genome. A total of 2,769 genes (72.05%) were assigned a putative function (by COG or NR blast). Two hundred ninety-eight genes were identified as ORFans (7.75%). The remaining 515 genes were annotated as hypothetical proteins (13, 40%). The distribution of genes into COGs functional categories is presented in Table 4. The properties and the statistics of the genome are summarized in Tables 4 and 5.

Table 4. Nucleotide content and gene count levels of the genome.

Attribute	Value	% of total[a]
Genome size (bp)	3,735,762	100
DNA coding region (bp)	3,239,020	86.70
DNA G+C content (bp)	1,045,424	27.98
Total genes	3,916	100
RNA genes	73	1.86
Protein-coding genes	3,843	98.14
Genes with function prediction	2,769	72.05
Genes assigned to COGs	2,849	74.13
Genes with peptide signals	410	10.67
Genes with transmembrane helices	1,016	26.44

[a]The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome

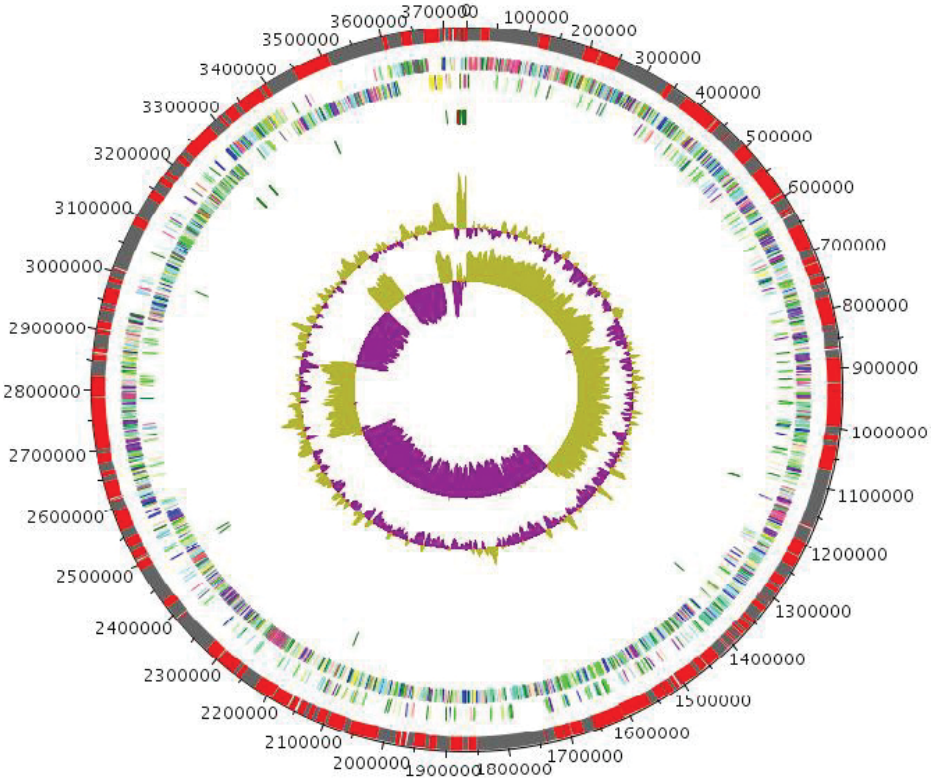


Figure 6. Graphical circular map of the chromosome. From the outside in, the outer two circles show open reading frames oriented in the forward and reverse directions (colored by COG categories), respectively. The third circle marks the rRNA gene operon (red) and tRNA genes (green). The fourth circle shows the G+C% content plot. The inner-most circle shows the GC skew, purple and olive indicating negative and positive values, respectively.

Comparison with the genomes from other *Clostridium* species

The genome sequence of *Clostridium sp.* is currently available for more than seventy-five *Clostridium* species. Here we compared the genome sequence of *C. dakarense* strain FF1^T with than those of *C. bartlettii*, *C. beijerinckii*, *C. cellulovorans*, *C. difficile*, *C. glycolicum*, *C. perfringens*, *C. saccharolyticum*, *C. senegalense*, and *C. thermocellum*.

The draft genome sequence of *C. dakarense* strain FF1^T is smaller than those of *C. cellulovorans*, *C. beijerinckii*, *C. senegalense*, *C. saccharolyticum*, *C. thermocellum*, *C. difficile*, *C. glycolicum* (3.73, 5.26, 6.0, 3.89, 4.66, 3.84, 4.3 and 3.99 Mb, respectively) but larger than those of *C. perfringens* and *C. bartletii* (3.26 and 2.97 Mb, respectively). The G+C content of *C. dakarense* is lower than those of *C. cellulovorans*, *C. beijerinckii*, *C. perfringens*, *C. saccharolyticum*, *C. thermocellum*, *C. difficile* (31.2, 29.9, 28.4, 45, 39 and 29.1%, respectively) but higher than those of *C. bartlettii*, *C. glycolicum* and *C. senegalense* (28.8, 28 and 26.8%, respectively). The gene content of *C. dakarense* is larger than those of *C. thermocellum*, *C. senegalense*, *C.*

perfringens, *C. glycolicum*, *C. bartlettii* (3,916, 3,173, 3,761, 2,876, 3,840 and 2,787, respectively) and smaller than those of *C. cellulovorans*, *C. beijerinckii*, *C. saccharolyticum* and *C. difficile*, (4,501, 5,243, 4,154 and 4,019, respectively). The ratio of genes per Mb of *C. dakarense* is larger to those of *C. cellulovorans*, *C. beijerinckii*, *C. senegalense*, *C. saccharolyticum*, *C. thermocellum*, *C. difficile*, *C. bartlettii*, *C. glycolicum* and *C. perfringens* (1,049, 856, 874, 966, 891, 826, 934, 938, 962 and 882, respectively).

The number of orthologous genes shared between *C. dakarense* and other compared *Clostridium* species has been summarized in Table 6. The average percentage of nucleotide sequence identity ranged from 62.05 to 74.5% among previously published *Clostridium* species, and from 61.94 to 75.7% between *C. dakarense* and other studied *Clostridium* species, thus confirming its new species status.

Table 5. Number of genes associated with the 25 general COG functional categories.

Code	Value	%age[a]	Description
J	171	4.45	Translation
A	0	0	RNA processing and modification
K	325	8.46	Transcription
L	158	4.11	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	34	0.88	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	111	2.89	Defense mechanisms
T	225	5.85	Signal transduction mechanisms
M	165	4.29	Cell wall/membrane biogenesis
N	58	1.51	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	45	1.17	Intracellular trafficking and secretion
O	95	2.47	Posttranslational modification, protein turnover, chaperones
C	194	5.04	Energy production and conversion
G	248	6.45	Carbohydrate transport and metabolism
E	248	6.45	Amino acid transport and metabolism
F	88	2.29	Nucleotide transport and metabolism
H	117	3.04	Coenzyme transport and metabolism
I	72	1.87	Lipid transport and metabolism
P	181	4.71	Inorganic ion transport and metabolism
Q	52	1.35	Secondary metabolites biosynthesis, transport and catabolism
R	386	10.04	General function prediction only
S	261	6.79	Function unknown
-	994	25.87	Not in COGs

[a]The total is based on the total number of protein coding genes in the annotated genome.

Table 6. Numbers of orthologous proteins shared between genomes (upper right)

	CDa	CC	CBe	CP	CSe	CSa	CT	CBa	CG	CDi
CDa	**3,808**	1,045	1,230	1,089	1,131	1,013	806	1,324	1,690	1,203
CC	68.22	**4,254**	1,490	1,163	1,181	1,057	967	871	1,038	1,021
CBe	68.84	70.36	**5,020**	1,300	1,289	1,207	968	989	1,204	1,129
CP	70.02	70.43	72.15	**2,660**	1,168	920	777	845	1,005	1,147
CSe	69.91	70.37	70.82	70.13	**3,704**	930	821	856	1,134	1,008
CSa	61.94	62.50	62.44	62.22	62.05	**4,154**	854	833	1,004	998
CT	64.49	64.84	64.56	64.78	64.53	63.83	**3,173**	713	840	952
CBa	74.98	68.22	68.84	69.46	69.52	62.15	64.73	**2,787**	1,517	1,303
CG	75.70	68.28	68.83	69.49	69.57	62.26	64.59	76.04	**3,840**	1,568
CDi	71.34	69.57	68.52	71.52	65.49	66.37	64.32	74.45	74.50	**3,798**

average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold). CDa: *C. dakarense*; CC: *C. cellulovorans*; CBe: *C. beijerinckii*; CP: *C. perfringens*; CSe: *C. senegalense*; CSa: *C. saccharolyticum*; CT: *C. thermocellum*; CBa: *C. bartlettii*; CG: *C. glycolicum*; CDi: *C. difficile*.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Clostridium dakarense* sp. nov. which contains strain FF1^T. This bacterium strain has been isolated from the fecal flora of a 4-months-old Senegalese child suffering from gastroenteritis.

Description of *Clostridium senegalense* sp. nov.

Clostridium dakarense (da.kar.e'n.se. L. gen. neutr. n. *dakarense*, pertaining to, or originating from Dakar, the capital of Senegal, where the type strain was isolated).

Colonies were 1.5 mm in diameter on blood-enriched Columbia agar and Chocolate agar + PolyViteX. Cells are rod-shaped with a mean diameter of 1.2 µm. Optimal growth is achieved anaerobically. No growth is observed in aerobic conditions. Growth occurs between 25-37°C, with optimal growth observed at 37°C, in medium 5% sheep blood-enriched Columbia agar. Cells stain Gram-positive, are endospore-forming, and motile. Catalase, oxidase, urease, indole and nitrate reduction activity are absent. Arginine dihydrolase, N-acetyl-β-glucosanimidase and pyroglutamic acid arylamidase activity are present. Cells are susceptible to amoxicillin, metronidazole, vancomycin, imipenem and rifampicin but resistant to trimethoprim/sulfamethoxazole.

The G+C content of the genome is 27.98%. The 16S rRNA gene sequence and whole-genome shotgun sequence of *C. dakarense* strain FF1^T (= CSUR P243 = DSM 27086) are deposited in GenBank under accession numbers KC517358 and CBTZ00000000, respectively. The type strain FF1^T (= CSUR P243 = DSM 27086) was isolated from the fecal flora of a 4-months-old child in Dakar, Senegal.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for automating the genomic annotation process. This study was funded by the Mediterranee-Infection Foundation.

References

1. Rossello-Mora R. DNA-DNA Reassociation Methods Applied to Microbial Taxonomy and Their Critical Evaluation. *In*: Stackebrandt E (ed), Molecular Identification, Systematics, and population Structure of Prokaryotes. Springer, Berlin, 2006, p. 23-50.
2. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; **33**:152-155.
3. Welker M, Moore ER. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst Appl Microbiol* 2011; **34**:2-11. PubMed http://dx.doi.org/10.1016/j.syapm.2010.11.013
4. Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 2010; **60**:249-266. PubMed http://dx.doi.org/10.1099/ijs.0.016949-0
5. Kokcha S, Mishra AK, Lagier JC, Million M, Leroy Q, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Bacillus timonensis* sp. nov. *Stand Genomic Sci* 2012; **6**:346-355. PubMed http://dx.doi.org/10.4056/sigs.2776064
6. Lagier JC, El Karkouri K, Nguyen TT, Armougom F, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. *Stand Genomic Sci* 2012; **6**:116-125. PubMed http://dx.doi.org/10.4056/sigs.2415480
7. Mishra AK, Gimenez G, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes senegalensis* sp. nov. *Stand Genomic Sci* 2012; **6**:304-314. http://dx.doi.org/10.4056/sigs.2625821
8. Lagier JC, Armougom F, Mishra AK, Ngyuen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. *Stand Genomic Sci* 2012; **6**:315-324. PubMed
9. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Clostridium senegalense* sp. nov. *Stand Genomic Sci* 2012; **6**:386-395. PubMed
10. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. *Stand Genomic Sci* 2012; **7**:1-11. PubMed http://dx.doi.org/10.4056/sigs.2956294
11. Mishra AK, Lagier JC, Rivet R, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Paenibacillus senegalensis* sp. nov. *Stand Genomic Sci* 2012; **7**:70-81. PubMed
12. Prazmowski A. "Untersuchung über die Entwickelungsgeschichte und Fermentwirking einiger Bakterien-Arten." Ph.D. Dissertation, University of Leipzig, Germany, 1880, p. 366-371.
13. List of prokaryotic names with standing with nomenclature. http://www.bacterio.net
14. Evaluations and Standards Laboratory. "Identification of *Clostridium* species", 2008 pp. 14.
15. Wells CL, Wilkins TD. (1996). *Clostridia:* Spore forming Anaerobic Bacilli In: *Baron's Medical Microbiology* (Baron S *et al.*, eds.) (4th ed.). University of Texas Medical Branch.
16. Keis S, Shaheen R, David T. Jones Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. *Int J Syst Evol Microbiol* 2001; **51**:2095-2103. PubMed http://dx.doi.org/10.1099/00207713-51-6-2095
17. Carnahan JE, Castle JE. Nitrogen Fixation. *Annu Rev Plant Physiol* 1963; **14**:125-136. http://dx.doi.org/10.1146/annurev.pp.14.060163.001013
18. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence(MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360
19. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archae, Bacteria, and Eukarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576
20. Gibbons NE, Murray RGE. Proposals Concerning the Higher Taxa of Bacteria. *Int J Syst Bacteriol*

1978; **28**:1-6. http://dx.doi.org/10.1099/00207713-28-1-1

21. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.

22. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), Bergey's Manual of Systematic Bacteriology, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.

23. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; **60**:469-472. http://dx.doi.org/10.1099/ijs.0.022855-0

24. Rainey FA. Class II. *Clostridia* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 736.

25. Skerman VBD, Sneath PHA. Approved list of bacterial names. *Int J Syst Bacteriol* 1980; **30**:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

26. Prévot AR. In: Hauderoy P, Ehringer G, Guillot G, Magrou. J., Prévot AR, Rosset D, Urbain A (eds), Dictionnaire des Bactéries Pathogènes, Second Edition, Masson et Cie, Paris, 1953, p. 1-692.

27. Pribram E. Klassification der Schizomyceten. Klassifikation der Schizomyceten (Bakterien), Franz Deuticke, Leipzig, 1933, p. 1-143.

28. Smith LDS, Hobbs G. Genus III. Clostridium Prazmowski 1880, 23. In: Buchanan RE, Gibbons NE (eds), Bergey's Manual of Determinative Bacteriology, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 551-572.

29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556

30. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; **33**:152-155.

31. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009; **49**:543-551. PubMed http://dx.doi.org/10.1086/600885

32. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

33. Prodigal. http://prodigal.ornl.gov

34. GenBank database. http://www.ncbi.nlm.nih.gov/genbank

35. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. PubMed

36. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. PubMed http://dx.doi.org/10.1093/nar/gkm160

37. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795. PubMed http://dx.doi.org/10.1016/j.jmb.2004.05.028

38. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; **305**:567-580. PubMed http://dx.doi.org/10.1006/jmbi.2000.4315

39. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics* 2000; **16**:944-945. PubMed http://dx.doi.org/10.1093/bioinformatics/16.10.944

40. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009; **25**:119-120. PubMed http://dx.doi.org/10.1093/bioinformatics/btn578

41. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004; **14**:1394-1403. PubMed http://dx.doi.org/10.1101/gr.2289704

42. Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011; **12**:124. PubMed http://dx.doi.org/10.1186/1471-2105-12-124

Article 5:

« High-quality genome sequence and description of Bacillus dielmoensis strain FF4^T sp. nov.»

Cheikh Ibrahima Lo[1,2,] Roshan Padhmanabhan[1,2], Oleg Mediannikov[1,2], Jérôme Terras[1,2],Catherine Robert[1,2], Ngor Faye[3], Didier Raoult[1,2,4], Pierre-Edouard Fournier[1,2]and Florence Fenollar[1,2]*

[1] Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de Médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France.

[2] Campus International UCAD-IRD, Dakar, Senegal

[3] Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann, Senegal.

[4] Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Publié dans:

Standards in Genomic Sciences.

http://www.standardsingenomics.com/content/10/1/41

Résumé de l'article 5: ***«High-quality finished genome sequence and description of Bacillus dielmoensis strain FF4^T sp. nov. »***

La souche FF4^T a été isolée de la flore cutanée d'une jeune femme Sénégalaise âgée de 16 ans et vivant à Dielmo au Sénégal. Le score observé par l'analyse de la spectrométrie de masse (MALDI-TOF) n'a pas permis d'identification (1.1 à 1.3). La séquence de l'ARN 16S ribosomique de cette souche a montré une similarité de 97,5% avec celle de *Bacillus fumarioli,* l'espèce la plus proche phylogénétiquement. Grâce à une étude polyphasique combinant des analyses phénotypiques et génomiques, la souche FF4^T a été décrite.

La souche FF4^T est une bactérie à Gram positif, aérobie et mobile. Elle ne forme pas de spores. Les colonies ont 2 mm de diamètre et sont de couleur blanche sur gélose Columbia enrichi au sang. Les bactéries sont en forme de bâtonnet avec un diamètre moyen de 0,6 µm (entre 0,5 et 0,8) et une longueur moyenne de 4,2 µm (entre 2,6 et 5,8). La croissance est observée en aérobie, mais une faible croissance est détectée en anaérobie. Une croissance est constatée entre 30 et 45°C, mais elle est optimale à 37°C. La catalase est présente mais l'oxydase est absente. Des réactions positives ont été obtenues pour l'esculine citrate de fer, l'estérase, l'estérase lipase, l'alcaline phosphatase, la naphthol-AS-BI-phosphohydrolase, l'acide phosphatase, la β-galactosidase, la β-glucuronidase, l'α-glucosidase et la β-glucosidase. *B. dielmoensis* est sensible *in vitro* à l'amoxicilline, amoxicilline-acide clavulanique, ceftriaxone, imipénème, ciprofloxacine, gentamicine, doxycycline, rifampicine, erythromycine mais résistante à la pénicilline, triméthoprime-sulfaméthoxazole et au métronidazole.

Le génome de la souche FF4^T est composé de 4 563 381 paires de base (1 chromosome mais pas de plasmide) avec 4 308 gènes codant pour des protéines et 157 gènes ARN (y compris 5 opérons ARNr). Son taux en G+C est de 40,8%.

Description de *Bacillus dielmoensis* sp. nov.

Sur la base des résultats, nous avons proposé la création de *Bacillus dielmoensis* sp. nov. (di.el.mo.en′sis. L. gen. masc. n. *dielmoensis* de Dielmo, le nom du village Sénégalais où la jeune femme, à partir de qui la souche a été isolée, vivait).

Les séquences de l'ARN 16S ribosomique et du génome ont été déposés à GenBank avec les numéros d'accession HG315676 et CCAD000000000, respectivement. La souche type FF4^T (= CSUR P3026 = DSM 27844) a été isolée de la flore cutanée d'une jeune femme en bonne santé vivant dans le village de Dielmo au Sénégal.

DOI 10.1186/s40793-015-0019-8

Standards in Genomic Sciences

SHORT GENOME REPORT Open Access

High-quality genome sequence and description of *Bacillus dielmoensis* strain FF4^T sp. nov.

Cheikh Ibrahima Lo[1,2], Roshan Padhmanabhan[1,2], Oleg Mediannikov[1,2], Jérôme Terras[1,2], Catherine Robert[1,2], Ngor Faye[3], Didier Raoult[1,2,4], Pierre-Edouard Fournier[1,2] and Florence Fenollar[1,2*]

Abstract

Strain FF4^T was isolated from the skin flora of a 16-year-old healthy Senegalese female. This strain exhibited a 16S rRNA sequence similarity of 97.5 % with *Bacillus fumarioli,* the phylogenetically closest species with standing in nomenclature and a poor MALDI-TOF-MS score (1.1 to 1.3) that does not allow any identification. Using a polyphasic study consisting of phenotypic and genomic analyses, strain FF4^T was Gram-positive, aerobic, rod-shaped, and exhibited a genome of 4,563,381 bp (1 chromosome but no plasmid) with a G + C content of 40.8 % that coded 4,308 protein-coding and 157 RNA genes (including 5 rRNA operons). On the basis of these data, we propose the creation of *Bacillus dielmoensis* sp. nov.

Keywords: *Bacillus dielmoensis*, Genome, Taxonogenomics, Culturomics

Introduction

The genus *Bacillus* (Cohn 1872) was created about 142 years ago [1]. Currently, the genus *Bacillus* comprised 281 species and 7 subspecies with validly published names [2]. Members of the genus *Bacillus* are environmental bacteria isolated most often from soil, food, fresh and sea water. Furthermore, they live rarely in human and animals in which they are either pathogens, such as *B. anthracis* (the causative agent of anthrax) [3, 4] and *B. cereus* (associated mainly with food poisoning) [4, 5], or saprophytes [4, 6]. Many species of the genus *Bacillus* are also isolated from different plants in which they are endophytes [7].

Recently, high throughput genome sequencing and mass spectrometric (MALDI-TOF MS) analyses of bacteria have given unprecedented access to an abundance of genetic and proteomic information [8–10]. Thus, a polyphasic approach is currently proposed to describe new bacterial taxa that includes their genome sequence, MALDI-TOF MS spectrum and major phenotypic characteristics such as Gram staining, culture, metabolic characteristics, habitat and if applicable, pathogenicity [9–11].

Bacillus dielmoensis strain FF4 (= CSUR P3026 = DSM 27844) is designated as the type strain of *B. dielmoensis.* This bacterium is a Gram-positive, non-spore-forming, aerobic and motile bacillus. This bacterium was isolated from the skin of a healthy Senegalese female as part of a "culturomics" study aiming at cultivating bacterial species from the skin flora [12]. Here, we present a summary classification and a set of features for *B. dielmoensis* sp. nov. strain FF4^T together with the description of the complete genomic sequencing and annotation. These characteristics support the circumscription of the species *B. dielmoensis.*

Organism information

Classification and features

A skin sample was collected with a swab from a healthy Senegalese volunteer living in Dielmo (a rural village in the Guinean-Sudanian area in Senegal) in December 2012 (Table 1). This 16-year-old healthy Senegalese female was included in a research project that was approved by the

* Correspondence: florence.fenollar@univ-amu.fr
[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
[2]Campus International UCAD-IRD, Dakar, Senegal
Full list of author information is available at the end of the article

Table 1 Classification and general features of *Bacillus dielmoensis* strain FF4^T [17]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain: *Bacteria*	TAS [31]
		Phylum: *Firmicutes*	TAS [32, 33]
		Class: *Firmibacteria*	TAS [34, 35]
		Order: *Bacillales*	TAS [31, 36]
		Family: *Bacillaceae*	TAS [31, 37]
		Genus: *Bacillus*	TAS [31, 38, 39]
		Species: *Bacillus dielmoensis*	IDA
		Type strain: FF4^T	IDA
	Gram stain	Positive	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non sporulating	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37 °C	IDA
	pH range; Optimum	7-7.6; 7.3	IDA
	Carbon source	Unknown	
MIGS-6	Habitat	Human skin	IDA
MIGS-6.3	Salinity	Not growth in BHI medium + 5 % NaCl	IDA
MIGS-22	Oxygen requirement	Aerobic	IDA
MIGS-15	Biotic relationship	Free living	IDA
MIGS-14	Pathogenicity	Unknown	
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection time	December 2012	IDA
MIGS-4.1	Latitude	13.7167	IDA
MIGS-4.2	Longitude	– 16.4167	IDA
MIGS-4.4	Altitude	45 m above sea level	IDA

[a]Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [40]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements

Ministry of Health of Senegal, the assembled village population and the National Ethics Committee of Senegal (CNERS, agreement numbers 09–022), as published elsewhere [13]. The strain FF4^T (Table 1) was isolated in December 2012 by cultivation on 5 % sheep blood enriched Columbia agar (BioMérieux, Marcy l'Etoile, France), under aerobic conditions. When the 16S rRNA of *B. dielmoensis* was compared to those of all species with validly published names listed in the list of prokaryotic names with standing in nomenclature from which we also retrieved the 16S rRNA sequences, *B. dielmoensis* strain FF4^T exhibited a 97.5 % nucleotide sequence similarity with *B. fumarioli* [14], which is the phylogenetically closest *Bacillus* species (Fig. 1). These values were lower than the 98.7 % 16S rRNA gene sequence threshold recommended by Meier-Kolthoff et al., 2013 to delineate a new species within genus *Bacillus* without carrying out DNA-DNA hybridization [15]. Different growth temperatures (25, 30, 37, 45 °C) were tested. Growth was observed at 30, 37, and 45 °C with the optimal growth obtained at 37 °C after 24 h of incubation. Colonies were 2 mm in diameter and white in color on blood-enriched Colombia agar. Growth of the strain was also tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux), and under aerobic conditions, with or without 5 % CO_2. Growth was observed in all the above mentioned conditions except in anaerobic conditions, where only weak growth was observed. Gram staining showed Gram-positive long rods (Fig. 2). A motility test was also positive. Cells grown on agar have a diameter

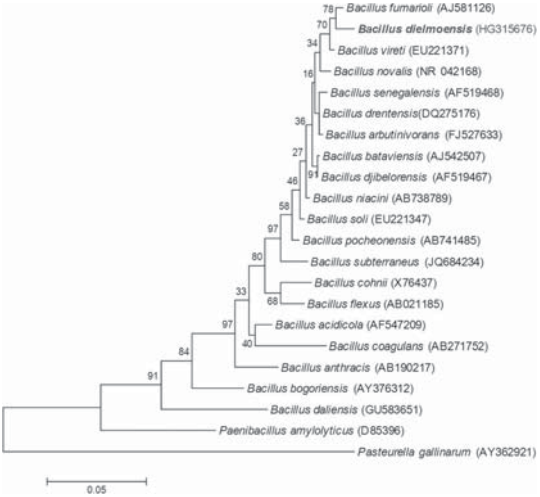


Fig. 1 Phylogenetic tree highlighting the position of *Bacillus dielmoensis* strain FF4^T relative to the most closely related type strains within the genus *Bacillus*. The strains and their corresponding GenBank accession numbers for 16S rRNA genes are (type = ^{T}) and in parenthesis, we indicated GA if the genome is available or GNA if the genome is not available in NCBI web site: *Bacillus fumarioli* strain CIP 106910^T (GNA), *B. dielmoensis* strain FF4^T, *B. vireti* LMG 21834 (GA : ALAN00000000), *B. novalis* strain IDA3307 (GNA), *B. senegalensis* strain RS8 (GNA), *B. drentensis* strain WN575 (GNA), *B. arbutinivorans* strain rif200874 (GNA), *B. bataviensis* strain LMG 21833 (GA: AJLS00000000), *B. djibelorensis* strain RS7 (GNA), *B. niacini* strain NBRC *15566*^{T} (GNA), *B. soli* strain NBRC 102451^T, *B. pocheonensis* strain GMC125 (GNA), *B. subterraneus* strain HWG-A11 (GNA), *B. cohnii* strain DSM 6307^T (GNA), *B. flexus* strain DSM 1320^T (GNA), *B. acidicola* strain 105-2^T (GNA), *B. coagulans* strain 2–6 (GA: CP002472), *B. anthracis* (GA: CP008854), *B. bogoriensis* strain ATCC BAA-922 (GA: JHYI00000000), *B. daliensis* strain DLS13 (GNA), *Paenibacillus amylolyticus* strain ATCC 9995^T and *Avibacterium gallinarum* strain NCTC 11188^T. Sequences were aligned using MUSCLE [41], and phylogenetic tree inferred using the Maximum Likelihood method with Kimura 2-parameter model from MEGA software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. The scale bar represents a rate of substitution per site of 0.05 was used as outgroup

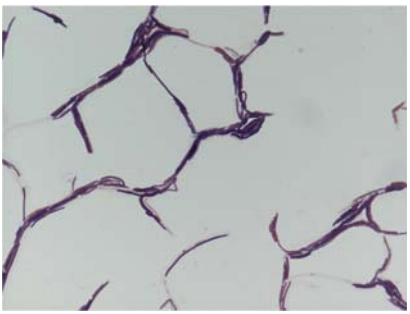


Fig. 2 Gram staining of *Bacillus dielmoensis* strain FF4^T

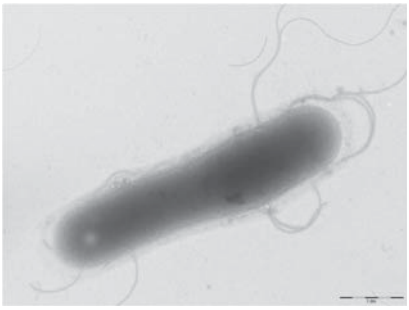


Fig. 3 Transmission electron microscopy of *Bacillus dielmoensis* strain FF4^T, using a Morgani 268D (Philips) at an operating voltage of 60 kV. The scale bar represents 1 μm

ranging from 0.5 to 0.8 μm and a length ranging from 2.6 to 5.8 μm as determined by negative staining transmission electron microscopy (Fig. 3).

Strain FF4^T exhibited catalase activity but not oxidase activity. Using the API 50 CH strip (BioMérieux), a positive reaction was observed only for esculin ferric citrate; all other reactions were negative including D-glucose, D-mannose, D-cellobiose, D-trehalose, D-raffinose, starch, D-lyxose, D-fucose, D-arabitol and potassium 2-KetoGluconate. Using the API ZYM strip (BioMérieux), positive reactions were obtained for esterase, esterase lipase, alkaline phosphatase, naphthol-AS-BI-phosphohydrolase, acid phosphatase, β-galactosidase, β-glucuronidase, α-glucosidase and β-glucosidase. No reaction was observed for α-galactosidase, α-chymotrypsin, trypsin, cystine arylamidase, valine arylamidase, leucine arylamidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Using the API 20E strip (BioMérieux), all the reactions were negative. *B. dielmoensis* is susceptible to amoxicillin, amoxicillin-clavulanic acid, ceftriaxone, imipenem, ciprofloxacin, gentamicin, doxycycline, rifampicin, erythromycin and vancomycin, but resistant to penicillin, trimethoprim-sulfamethoxazole and metronidazole. When compared with representative species from the genus *Bacillus*, *B. dielmoensis* strain FF4^T exhibited the phenotypic differences detailed in Additional file 1: Table S1.

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was performed using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany), as previously reported [16]. The scores previously established by Bruker Daltonics allowing validating or not the identification of species compared to the database of the instrument were applied. Briefly, a score ≥ 2 with a species with a validly published name provided allows the identification at the species level; a score ≥ 1.7 and < 2 allows the identification at the genus level; and a score < 1.7 does not allow any identification. We performed 12 distinct deposits from 12 isolated colonies of strain FF4^T. Two microliters of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50 % acetonitrile and 2.5 % trifluoroacetic-acid were distributed on each smear and submitted at air drying for five minutes. Then, the spectra from the 12 different colonies were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against the main

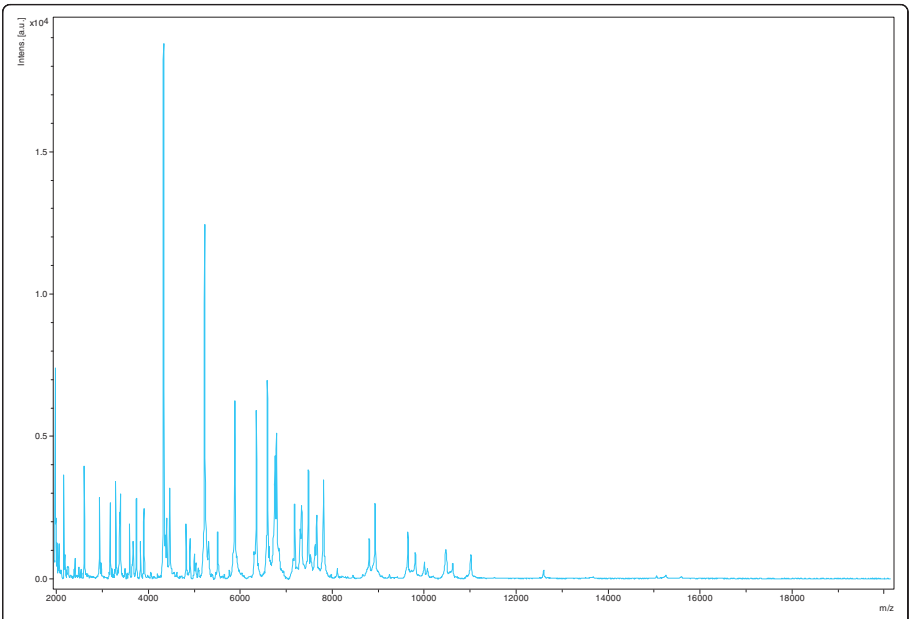


Fig. 4 Reference mass spectrum from *Bacillus dielmoensis* strain FF4^T. Spectra from 12 individual colonies were compared and a reference spectrum was generated

spectra of 6,252 bacterial spectra including 199 spectra from 104 *Bacillus* species. Scores ranged from 1.1 to 1.3 were obtained for the strain FF4^T, suggesting that this isolate was not a member of any known species. The reference mass spectrum from strain FF4^T was incremented in our database (Fig. 4). The gel view highlighted spectrum differences with other *Bacillus* species (Fig. 5).

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its 16S rRNA similarity, phylogenetic position and phenotypic differences with other members of the genus *Bacillus*, which support that *Bacillus dielmoensis* strain FF4^T likely represents a new bacterial species. Besides, this strain is part of a study aiming to characterize the skin flora of healthy Senegalese people. Currently, there are more of 270 sequenced genomes of *Bacillus* species [2]. The strain FF4^T is the first genome of *B. dielmoensis* sp. nov. GenBank accession number is CCAD000000000. It consists of 75 contigs. Table 2 shows the project information and its association with MIGS version 2.0 compliance [17]. Associated MIGS records are detailed in Additional file 2: Table S2.

Table 2 Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	One 454 paired-end 3-kb library
MIGS-29	Sequencing platforms	454 GS FLX Titanium
MIGS-31.2	Fold coverage	61x
MIGS-30	Assemblers	Newbler version 2.5.3
MIGS-32	Gene calling method	Prodigal
	Locus Tag	Not determinated
	GenBank ID	CCAD000000000
	GenBank Date of Release	March 12, 2014
	GOLD ID	Gp0101145
	BIOPROJECT	PRJEB4276
MIG-13	Source Material Identifier	DSM 27844
	Project relevance	Study of human skin flora

Growth conditions and genomic DNA preparation

Bacillus dielmoensis strain FF4^T (= CSUR P3026 = DSM 27844) was grown aerobically on 5 % sheep blood-enriched Columbia agar (BioMérieux) at 37 °C. Bacteria growing in four Petri dishes were suspended in 5x100 µL of TE buffer. Then, 150 µL of this suspension were diluted in 350 µL TE buffer 10X, 25 µL proteinase K and 50 µL sodium dodecyl sulfate (SDS) for lysis treatment. This preparation was incubated overnight at 56 °C. DNA was washed 3 times using UltraPure™ Phenol:Chloroform: Isoamyl Alcohol (25:24:1, v/v) (Thermo Fisher Scientific Inc, Waltham, USA) and was precipitated with ethanol at −20 °C during overnight. Following centrifugation, DNA was suspended in 65 µL EB buffer. The genomic DNA concentration was measured at 43.96 ng/µL using the Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, USA).

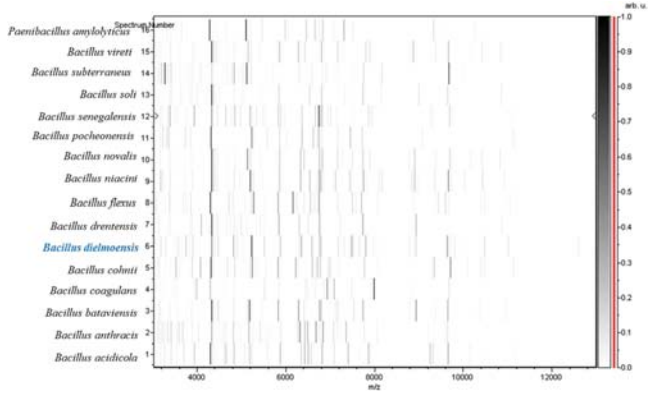


Fig. 5 Gel view comparing *Bacillus dielmoensis* strain FF4^T to other members of the genus *Bacillus*. The gel view displays the raw spectra of all loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Gray scale scheme code. The color bar and the right y-axis indicate the relation between the color a peak is displayed with and the peak intensity in arbitrary units. Displayed species are indicated on the left

Genome sequencing and assembly

Genomic DNA of *Bacillus dielmoensis* was sequenced on the MiSeq Technology (Illumina Inc, San Diego, CA, USA) with the 2 applications: paired-end and mate-pair. The paired-end and the mate-pair strategies were barcoded in order to be mixed respectively with 10 others genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and 11 others projects with the Nextera Mate-Pair sample prep kit (Illumina).

Genomic DNA was diluted to 1ng/μL to prepare the paired-end library. The "tagmentation" step fragmented and tagged the DNA with an optimal size distribution at 1.6 kb. Then, limited cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter Inc, Fullerton, CA, USA), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in single 39-h run in 2 × 250-bp.

A total of 3.89 Gb sequence was obtained from a 416 K/mm2 cluster density with a cluster passing quality control filters of 95.4 % (7,899,000 clusters). *B. dielmoensis* strain FF4^T showed an index representation of 4.95 % within the run and presented 373,015 reads filtered according to the read qualities.

The mate-pair library was prepared with 1 μg of genomic DNA using the Nextera mate-pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1.5 kb up to 10 kb with an optimal size at 5 kb. No size selection was performed and 600 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA) at 586 bp. The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 10 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-h run in a 2×250-bp.

Global information of 3.2 Gb was obtained from a 690 K/mm2 cluster density with a cluster passing quality control filters of 95.4 % (13,264,000 clusters). *B. dielmoensis* strain FF4^T shown an index representation of 8.02 % within the run and presented 1,014,931 reads filtered according to the read qualities.

Genome annotation

Open Reading Frame prediction of the *B. dielmoensis* FF4^T genome was performed using Prodigal [18] with default parameters. We excluded the predicted ORFs if they spanned a sequencing gap region. Functional assessment of protein sequences was carried out by comparing them with sequences in the GenBank [19] and Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAs, rRNAs, signal peptides and transmembrane helices were identified using tRNAscan-SE 1.21 [20], RNAmmer [21], SignalP [22] and TMHMM [23], respectively. Artemis [24] was used for data management whereas DNA Plotter [25] was used for visualization of genomic features. In house perl and bash scripts were used to automate these routine tasks. ORFans were sequences which have no homology in a given database *i.e.* in a non-

Table 3 Orthologous gene comparison of *Bacillus dielmoensis* strain FF4^T with other closely related species. Bold numbers indicate the number of genes from each genome

	Bacillus dielmoensis strain FF4^T	*Bacillus bataviensis* strain LMG 21833	*Bacillus coagulans* strain 2-6	*Bacillus coagulans* strain 36D1	*Bacillus isronensis* strain B3W22	*Lysinibacillus sphaericus* strain C3-41
Bacillus dielmoensis strain FF4^T	**4,308**					
Bacillus bataviensis strain LMG 21833	1,888	**5,207**				
Bacillus coagulans 2-6	1,517	1,617	**2,971**			
Bacillus coagulans strain 36D1	1,631	1,737	1,824	**3,289**		
Bacillus isronensis strain B3W22	1,545	1,681	1,332	1,434	**3,883**	
Lysinibacillus sphaericus strain C3-41	1,512	1,669	1,321	1,413	1,965	**4,584**

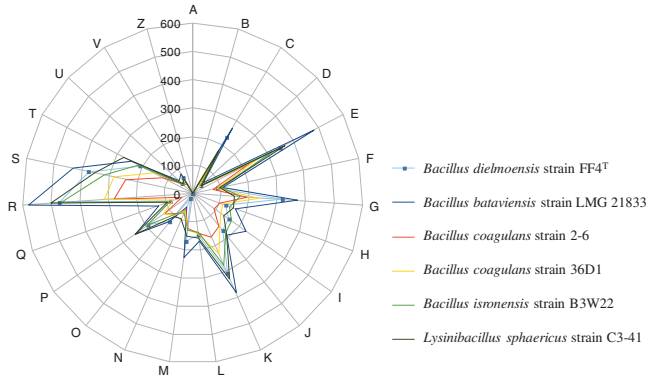


Fig. 6 Distribution of functional classes of predicted genes of *B. dielmoensis* strain FF4^T along with other *Bacillus* genomes according to the clusters of orthologous groups of proteins

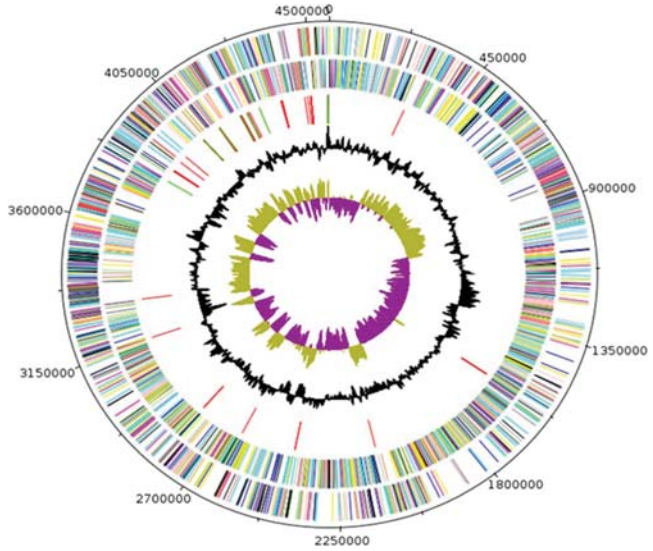


Fig. 7 Graphical circular map of the *Bacillus dielmoensis* strain FF4^T chromosome. From the outside in, the outer two circles shows open reading frames oriented in the forward (colored by COG categories) and reverse (colored by COG categories) direction, respectively. The third circle marks the rRNA gene operon (red) and tRNA genes (green). The fourth circle shows the G + C% content plot. The inner-most circle shows GC skew, purple indicating negative values whereas olive for positive values

redundant (nr) or identified if their BLASTP E-value was lower than 1e-03 for alignment lengths greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E-value of 1e-05. PHAST was used to identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids [26].

To estimate the nucleotide sequence similarity at the genome level between *B. dielmoensis* and other members of the genus *Bacillus* (Table 3, Fig. 6), orthologous proteins were detected using the Proteinortho software [27] (with the parameters: e-value 1e-5, 30 % percentage of identity, 50 % coverage and algebraic connectivity of 50 %) and genomes compared two by two. After fetching the corresponding nucleotide sequences of orthologous proteins for each pair of genomes, we determined the mean percentage of nucleotide sequence identity using the Needleman-Wunsch global alignment algorithm. The script was created to calculate the average genomic identity of orthologous gene sequences (AGIOS) between genomes using the MAGi software (Marseille Average genomic identity). The script created to calculate AGIOS values was named MAGi (Marseille Average genomic identity) and is written in perl and bioperl modules. GGDC analysis was also performed using the GGDC web server as previously reported [28, 29].

Table 4 Nucleotide content and gene count levels of the genome

Attribute		Genome (total)
	Value	% of total[a]
Genome size (bp)	4,563,381	100
DNA coding (bp)	3,902,509	85.5
DNA G + C (bp)	1,864,870	40.8
DNA scaffolds	nd[b]	
Total genes	4,465	100
Protein-coding genes	4,308	96.4
RNA genes	157	
Pseudo genes	51	1.18
Genes in internal clusters	208	4.82
Genes with function prediction	2,847	66.0
Genes assigned to COGs	3,216	74.6
Genes with Pfam domains	3,235	75.4
Genes with peptide signals	137	3.18
Genes with transmembrane helices	1,189	27.5
CRISPR repeats	3	

[a]The total is based on either the size of genome in base pairs or the total number of protein coding genes in the annotated genome
[b]nd: not determined

Table 5 Number of genes associated with the 25 general COG functional categories

Code	Value	% of total[a]	Description
J	155	3.60	Translation
A	0	0.00	RNA processing and modification
K	216	5.01	Transcription
L	126	2.92	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	33	0.77	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	70	1.62	Defense mechanisms
T	125	2.90	Signal transduction mechanisms
M	152	3.52	Cell wall/membrane biogenesis
N	0	0.00	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	24	0.55	Intracellular trafficking and secretion
O	98	2.27	Posttranslational modification, protein turnover, chaperones
C	197	4.57	Energy production and conversion
G	233	5.40	Carbohydrate transport and metabolism
E	260	6.03	Amino acid transport and metabolism
F	70	1.62	Nucleotide transport and metabolism
H	86	1.99	Coenzyme transport and metabolism
I	100	2.32	Lipid transport and metabolism
P	147	3.41	Inorganic ion transport and metabolism
Q	26	0.60	Secondary metabolites biosynthesis, transport and catabolism
R	381	8.84	General function prediction only
S	348	8.91	Function unknown
-	369	8.56	Not in COGs

[a]The total is based on the total number of protein coding genes in the annotated genome

Genome properties

The genome of *B. dielmoensis* strain FF4^T is 4,563,381 bp long (1 chromosome but no plasmid) with a 40.8 % G + C content (Fig. 7). Of the 4,465 predicted genes, 4,308 were protein-coding genes and 157 were RNAs. A total of 3,216 genes (74.6 %) were assigned to COGs. A total of 137 genes were annotated as genes with peptide signals. The properties and the statistics of the genome are presented in Table 4. The distribution of genes into COGs functional categories is presented in Table 5.

Insights from the genome sequence

Today there are more than 277 sequenced genomes of *Bacillus* species (finished and draft) available in genomic databases [8]. Here, we have compared *B. dielmoensis*

Table 6 Comparison of *Bacillus dielmoensis* strain FF4^T with genomes of other *Bacillus* species and those of *Lysinibacillus sphaericus*

Microrganisms	Accession number	Number of proteins	G + C%	Genome size (Mb)
Bacillus dielmoensis strain FF4^T	CCAD000000000	4,308	40.8	4,56
Bacillus bataviensis LMG 21833	AJLS00000000	5,207	39.6	5,37
Bacillus coagulans 2-6	NC_015634	2,971	47.3	3,07
Bacillus coagulans 36D1	NC_016023	3,289	46.5	3,55
Bacillus isronensis B3W22	AMCK00000000	3,883	38.8	4,02
Lysinibacillus sphaericus C3-41	NC_010382	4,584	37.1	4,82

genome sequences against other members of genus *Bacillus* including *B. coagulans* strain 2–6, *B. coagulans* strain 36D1, *B. bataviensis* strain LMG 21833, *B. isronensis* strain B3W22, and *Lysinibacillus sphaericus* strain C3-41. The Table 6 shows a comparison of genome size, G + C% content, and number of proteins for selected *Bacillus* genomes for taxonogenomic study.

Bacillus dielmoensis strain FF4^T has a G + C content (40.8) lower than those of *Bacillus coagulans* 2–6 and 36D1 (47.3 and 46.5, respectively) but higher than those of *B. bataviensis* LMG 21833, *B. isronensis* B3W22 and *L. sphaericus* C3-41 (39.6, 38.8 and 37.1, respectively). As it has been suggested in the literature that the G + C content deviation is at most 1 % within species, these data are an additional argument for the creation of a new taxon [30].

Figure 6 shows the comparison of gene distribution into COG categories of *B. dielmoensis* with other finished genomes mentioned above. Table 3 presents the numbers of orthologous genes between genome pairs. Table 7 summarizes the AGIOS and dDDH values between the studied genomes. The AGIOS values ranged from 63.25 to 73.22 % at the interspecies level, between *B. dielmoensis* and other species, but was of 95.94 % at the intraspecies level, between the two *B. coagulans* strains. We obtained similar results using the GGDC software, as dDDH values ranged from 0.1057 to 0.2321 between studied species, and was 0.0505 between *B. coagulans* strains. These values confirm the status of *B. dielmoensis* as a new species.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses (taxonogenomics), we formally propose the creation of *Bacillus dielmoensis* sp. nov. that contains the strain FF4^T as type strain. The strain was isolated from the skin of a healthy Senegalese 16-year-old female living in Dielmo, Senegal.

Description of *Bacillus* dielmoensis sp. nov.

Bacillus dielmoensis (di.el.mo.en'sis. L. gen. masc. n. *dielmoensis* of Dielmo, the name of the Senegalese village where the female, from whom strain FF4^T was cultivated).

Bacillus dielmoensis is an aerobic Gram-positive bacterium, non-endospore forming and motile. Colonies are 2 mm in diameter and white in color on blood-enriched Colombia agar. Cells are rod-shaped with a mean diameter of 0.6 μm (range 0.5 to 0.8) and a mean length of 4.2 μm (range 2.6 to 5.8). Optimal growth is observed aerobically, weak growth occurs under anaerobic conditions. Growth occurs between 30 and 45 °C, with optimal growth occurring at 37 °C. A catalase activity is present but not oxidase activity. A positive reaction is obtained only for esculin ferric citrate. Positive reactions are observed for esterase, esterase lipase, alkaline phosphatase, naphthol-AS-BI-phosphohydrolase, acid phosphatase, β-

Table 7 dDDH values (upper right) and AGIOS values (lower left) obtained by comparison of all studied genomes

	Bacillus dielmoensis strain FF4T	*Bacillus bataviensis* strain LMG 21833	*Bacillus coagulans* strain 2-6	*Bacillus coagulans* strain 36D1	*Bacillus isronensis* strain B3W22	*Lysinibacillus sphaericus* strain C3-41
Bacillus dielmoensis strain FF4T		0.2321	0.1385	0.1069	0.1866	0.1553
Bacillus bataviensis strain LMG 21833	73.22		0.1658	0.1395	0.207	0.1554
Bacillus coagulans strain 2-6	64.84	64.71		0.0505	0.1316	0.1057
Bacillus coagulans strain 36D1	64.62	64.49	95.94		0.1571	0.107
Bacillus isronensis strain B3W22	63.25	63.31	62.21	62.11		0.1981
Lysinibacillus sphaericus strain C3-41	63.5	63.61	61.78	61.76	69.18	

galactosidase, β-glucuronidase, α-glucosidase and β-glucosidase. *B. dielmoensis* is susceptible to amoxicillin, amoxicillin-clavulanic acid, ceftriaxone, imipenem, ciprofloxacin, gentamicin, doxycycline, rifampicin, erythromycin and vancomycin, but resistant to penicillin, trimethoprim-sulfamethoxazole and metronidazole.

The G + C content of the genome is 40.8 %. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers HG315676 and CCAD000000000, respectively. The type strain FF4^T (= CSUR P3026 = DSM 27844) was isolated from the skin of a healthy female in Dielmo, Senegal.

Additional files

Additional file 1: Table S1. Differential phenotypic characteristics of *Bacillus dielmoensis* strain FF4^T and other *Bacillus* strains [4, 7 14, 42–44].

Additional file 2: Table S2. Associated MIGS record.

Abbreviations

CSUR: Collection de Souches de l'Unité des Rickettsies; DSM: Deutsche Sammlung von Mikroorganismen; CNERS: National Ethics Committee of Senegal; BHI: Brain Heart Infusion; MALDI-TOF MS: Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry; TE buffer: Tris-EDTA buffer; SDS: sodium dodecyl sulfate; MAGi: Marseille Average genomic identity; AGIOS: Average Genomic Identity of Orthologous Gene Sequences; GGDC: Genome-to-genome distance calculator; dDDH: Digital DNA-DNA hybridization.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CIL performed the phenotypic characterization of the bacterium and drafted the manuscript. RP performed the genomic analyses and drafted the manuscript. OM participated in its design and helped to draft the manuscript. CR performed the genomic sequencing and helped to draft the manuscript. JT helped to perform the phenotypic characterization of the bacterium and to draft the manuscript. NF participated in its design and helped to draft the manuscript. DR conceived the study and helped to draft the manuscript. PEF and FF conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This study was funded by the Méditerranee Infection Foundation.

Author details

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France. [2]Campus International UCAD-IRD, Dakar, Senegal. [3]Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann, Senegal. [4]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia.

Received: 4 July 2014 Accepted: 13 May 2015
Published online: 22 July 2015

References

1. Cohn F. Untersuchungen über Bakterien. Beitrage zur Biologie der Pflanzen Heft. 1872;1:127–224.
2. Abstract for the genus *Bacillus*. Names for Life, LLC. Retrieved September 22, 2013. Pubmed http://doi.namesforlife.com/10.1601/nm.4857.
3. Jernigan JA, Stephens DS, Ashford DA, Omenaca C, Topiel MS, Galbraith M, et al. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg Infect Dis. 2001;7:933–44. PubMed http://dx.doi.org/10.3201/eid0706.010604.
4. Keita MB, Diene SM, Robert C, Raoult D, Fournier PE, Bittar F. Non-contiguous finished genome sequence and description of *Bacillus massiliogorillae* sp. nov. Stand Genomic Sci. 2013;9:93–105. PubMed http://dx.doi.org/10.4056/sigs.4388124.
5. Bottone EJ. *Bacillus cereus*, a volatile human pathogen. Clin Microbiol Rev. 2010;23:382–98. PubMed http://dx.doi.org/10.1128/CMR.00073-09.
6. Mandell GL, Bennett JE, Dolin R. Principles and Practice of Infectious Diseases. Elsevier; Philadelphia, USA.: 2010. p. 4320.
7. Zhang YZ, Chen WF, Li M, Sui XH, Liu HC, Zhang XX, et al. *Bacillus endoradicis* sp. nov., an endophytic bacterium isolated from soybean root. Int J Syst Evol Microbiol. 2012;62:359–63. Pubmed http://dx.doi.org/10.1099/ijs.0.028936-0.
8. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, et al. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. 2012;40:D571–9. Pubmed http://dx.doi.org/10.1093/nar/gkr1100.
9. Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol. 2014;64:384–91. Pubmed http://dx.doi.org/10.1099/ijs.0.057091-0.
10. Sentausa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect. 2013;19:790–5. Pubmed http://dx.doi.org/10.1111/1469-0691.12181.
11. Lo CI, Mishra AK, Padhmanabhan R, Samb B, Sow AG, Robert C, et al. Non-contiguous finished genome sequence and description of *Clostridium dakarense* sp. nov. Stand Genomic Sci. 2013;30:14–27. Pubmed http://dx.doi.org/10.4056/sigs.4097825.
12. Lagier JC, Million M, Hugon P, Armougom F, Raoult D. Human gut microbiota: repertoire and variations. Front Cell Infect Microbiol. 2012;2:136. Pubmed http://dx.doi.org/10.3389/fcimb.2012.00136.
13. Trape JF, Tall A, Diagne N, Ndiath O, Ly AB, Faye J, et al. Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. Lancet Infect Dis. 2011;11:925–32. PubMed http://dx.doi.org/10.1016/S1473-3099(11)70194-3.
14. Logan NA, Lebbe L, Hoste B, Goris J, Forsyth G, Heyndrickx M, et al. Aerobic endospore-forming bacteria from geothermal environments in northern Victoria Land, Antarctica, and Candlemas Island, South Sandwich archipelago, with the proposal of *Bacillus fumarioli* sp. nov. Int J Syst Evol Microbiol. 2000;50:1741–53. Pubmed http://dx.doi.org/10.1099/00207713-50-5-1741.
15. Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol. 2013;195(6):413–8. Pubmed http://dx.doi.org/10.1007/s00203-013-0888-4.
16. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009;49:543–51. PubMed http://dx.doi.org/10.1086/600885.
17. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:541–7. PubMed http://dx.doi.org/10.1038/nbt1360.
18. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
19. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2012;40:48–53. PubMed http://dx.doi.org/10.1093/nar/gkr1202.
20. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
21. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8. PubMed http://dx.doi.org/10.1093/nar/gkm160.
22. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95. PubMed http://dx.doi.org/10.1016/j.jmb.2004.05.028.
23. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80. PubMed http://dx.doi.org/10.1006/jmbi.2000.4315.
24. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics. 2000;16:944–5. PubMed http://dx.doi.org/10.1093/bioinformatics/16.10.944.

25. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics. 2009;25:119–20. PubMed http://dx.doi.org/10.1093/bioinformatics/btn578.
26. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39:347–52. PubMed http://dx.doi.org/10.1093/nar/gkr485.
27. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;12:124. PubMed http://dx.doi.org/10.1186/1471-2105-12-124.
28. Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci. 2010;2:142–8.
29. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14:60.
30. Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol. 2014;64:352–6.
31. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eukarya*. Proc Natl Acad Sci U S A. 1990;87:4576–9. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576.
32. Skerman VBD, Sneath PHA. Approved List of Bacterial Names. Int J Syst Bacteriol. 1980;30:225–420. PubMed http://dx.doi.org/10.1099/00207713-30-1-225.
33. Garrity GM, Holt J. The road map to the Manual. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's Manual of Systematic Bacteriology, vol. 1. 2nd ed. New York: Springer; 2001. p. 119–69.
34. List no. 132. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol. 2010;60:469–72. PubMed http://dx.doi.org/10.1099/ijs.0.022855-0.
35. Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB, editors. Bergey's manual of systematic bacteriology, vol. 3. 2nd ed. New York: Springer; 2009. p. 19–20.
36. Prevot AR. Dictionnaire des bactéries pathogènes. In: Hauduroy P, Ehringer G, Guillot G, Magrou J, Prevot AR, Rosset, Urbain A, editors. Paris: Masson; 1953. p. 1–692.
37. Fischer A. Untersuchungen über bakterien. Jahrbücher für Wissenschaftliche Botanik. 1895;27:1–163.
38. Gibson T, Gordon RE. Genus I: *Bacillus* Cohn 1872, 174; Nom. gen. cons. Nomencl. Comm. Intern. Soc. Microbiol. 1937, 28; Opin. A. Jud. Comm. 1955, 39. In: Buchanan RE, Gibbons NE, editors. Bergey's Manual of Determinative Bacteriology. 8th ed. Baltimore: The Williams and Wilkins Co; 1974. p. 529–50.
39. Mathews WC, Caperna J, Toerner JG, Barber RE, Morgenstern H. Neutropenia is a risk factor for Gram-negative *Bacillus* bacteremia in human immunodeficiency virus-infected patients: results of a nested case–control study. Am J Epidemiol. 1998;148:1175–83. PubMed http://dx.doi.org/10.1093/oxfordjournals.aje.a009606.
40. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9. PubMed http://dx.doi.org/10.1038/75556.
41. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
42. Heyrman J, Vanparys B, Logan NA, Balcaen A, Rodríguez-Díaz M, Felske A, et al. *Bacillus novalis* sp. nov., *Bacillus vireti* sp. nov., *Bacillus soli* sp. nov., *Bacillus bataviensis* sp. nov. and *Bacillus drentensis* sp. nov., from the Drentse A grasslands. Int J Syst Evol Microbiol. 2004;54(Pt 1):47–57.
43. Nedashkovskaya OI, Van Trappen S, Frolova GM, De Vos P. *Bacillus berkeleyi* sp. nov., isolated from the sea urchin *Strongylocentrotus intermedius*. Arch Microbiol. 2012;194:215–21. PubMed http://dx.doi.org/10.1007/s00203-011-0771-0.
44. Jeroen H, An B, Marina RD, Niall AL, Jean S, Paul DV. *Bacillus decolorationis* sp. nov., isolated from biodeteriorated parts of the mural paintings at the Servilia tomb (Roman necropolis of Carmona, Spain) and the Saint-Catherine chapel (Castle Herberstein, Austria). Int J Syst Evol Microbiol. 2003;53:459–63. PubMed http://dx.doi.org/10.1099/ijs.0.02452-0.

Article 6:

«Non-contiguous finished genome sequence and description of Weeksella massiliensis sp. nov. »

Senthil Alias Sankar[1]*, **Cheikh Ibrahima Lo**[1,3]*, Bécaye Fall[2], Bissoume Sambe-Ba[2], Oleg Mediannikov[1,3], Ibrahima Diallo[2], Noémie Labas[1], Ngor Faye[4], Boubacar Wade[2], Didier Raoult[1,3,5], Pierre-Edouard Fournier[1] and Florence Fenollar[1,3]

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de Médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France Tel: +33 491 385 517

[2]Hôpital Principal, Dakar, Senegal

[3]Campus international UCAD-IRD, Dakar, Senegal

[4]Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Dakar, Senegal

[5]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Publié dans:

New Microbes and New Infections

November 2015, Volume 8, Pages 89-98

http://dx.doi.org/10.1016/j.nmni.2015.09.013

Résumé de l'article 6: ***«Non-contiguous finished genome sequence and description of Weeksella massiliensis sp. nov. »***

La souche $FF8^T$ a été isolée à l'Hôpital Principal de Dakar, Sénégal, dans l'urine d'un homme de 65 ans souffrant de cystite aiguë. L'analyse par spectrométrie de masse MALDI-TOF de la souche $FF8^T$ n'a pas permis d'identification (les scores étant compris entre 1.32 et 1.56). La souche $FF8^T$ a montré une similarité de séquence de l'ARN 16S ribosomique de 98,38% avec *Weeksella virosa*, l'espèce officielle la plus proche phylogénétiquement.

Nous avons réalisé une étude polyphasique reposant sur des analyses phénotypiques et génomiques. La souche $FF8^T$ est une bactérie à Gram négatif, aérobie, membre de la famille des *Flavobacteriaceae*. Les bactéries sont en forme de bâtonnet avec un diamètre moyen de 0,3 µm (entre 0,2-0,5 µm) et une longueur moyenne de 1,5 µm (entre 0,8-2,1µm). Elles sont immobiles. Sur gélose au sang Columbia enrichi avec 5% de sang de mouton, les colonies ont un diamètre de 2 mm ; elles sont opaques et légèrement jaune avec une surface lisse. La catalase et l'oxydase sont positives. Des réactions positives ont été aussi observées pour la phosphatase alcaline, l'estérase, l'estérase-lipase, la leucine arylamidase, l'acide phosphatase et le Naphthol-AS-BI-phosphohydrolase.

La souche $FF8^T$ est sensible *in vitro* à la ceftriaxone, l'amoxicilline/acide clavulanique, la pénicilline, l'imipénème, la gentamicine, et la doxycycline mais résistante à la nitrofurantoïne, la vancomycine, le triméthoprime-sulfaméthoxazole et le métronidazole.

Le génome comporte 2 562 781 paires de bases (1 chromosome mais pas de plasmide) et possède 2 390 gènes codant pour des protéines et 56 gènes ARN, y compris un opéron ARNr. Le taux en G+C est de 35,9%.

Description de *Weeksella massiliensis* sp. nov

Nous proposons formellement la création de *Weeksella massiliensis* sp. nov. qui contient la souche $FF8^T$ comme souche type. *Weeksella massiliensis* (mas.il.i.en'sis. L. gen. fem. n. *massiliensis*, de *Massilia*, le nom Latin de Marseille, France, où la souche $FF8^T$ a été caractérisée). Les séquences de l'ARN 16S ribosomique et du génome ont été déposées dans GenBank avec les numéros d'accession respectifs HG931340 et CCMH00000000. La souche type $FF8^T$ (= CSUR P860 = DSM 28259) a été isolée d'un prélèvement urinaire d'un homme de 65 ans souffrant de cystite aiguë à l'hôpital Principal de Dakar, Sénégal.

Noncontiguous finished genome sequence and description of *Necropsobacter massiliensis* sp. nov.

C. I. Lo[1,3], R. Padhamanabhan[1], B. Fall[2], B. Sambe-Ba[2], O. Mediannikov[1,3], T.-T. Nguyen[1], E. Prudent[1], N. Faye[4], B. Wade[2], D. Raoult[1,3,5], P.-E. Fournier[1] and F. Fenollar[1,3]

1) *Faculté de médecine, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Aix-Marseille Université, Marseille, France,* 2) *Hôpital Principal,* 3) *Campus International UCAD-IRD,* 4) *Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Dakar, Senegal and* 5) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Strain FF6^T was isolated from the cervical abscess of a 4-year-old Senegalese boy, in Dakar, Senegal. MALDI-TOF MS did not provide any identification. This strain exhibited a 95.17% 16S rRNA sequence identity with *Necropsobacter rosorum*. Using a polyphasic study including phenotypic and genomic analyses, strain FF6^T was an aero-anaerobic Gram-negative cocobacillus, oxidase positive, and exhibited a genome of 2,493,927 bp (1 chromosome but no plasmid) with a G+C content of 46.2% that coded 2,309 protein-coding and 53 RNA genes. On the basis of these data, we propose the creation of *Necropsobacter massiliensis* sp. nov.

Keywords: Culturomics, genome, *Necropsobacter massiliensis*, Senegal, taxono-genomics

Original Submission: 15 June 2015; **Revised Submission:** 4 September 2015; **Accepted:** 7 September 2015

Article published online: 16 September 2015

Corresponding author: F. Fenollar, Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France

E-mail: florence.fenollar@univ-amu.fr

Introduction

The genus *Necropsobacter* (Christensen *et al.* 2011) was first described in 2011 [1]. At this time, there is only one species with a validly published name [2]. In 2013, five clinical cases of bacteraemia associated with *Necropsobacter rosorum* were reported [3]. Members of the genus *Necropsobacter* were previously associated with the SP group that comprised mainly strains isolated from rabbits, rodents and humans [3]. Because *Necropsobacter rosorum* was the only described species in this genus with no genome available, we first sequenced its genome for genomic comparison [4]. *Necropsobacter massiliensis* strain FF6^T (= Collection de souches de l'Unité des Rickettsies (CSUR) P3511 = Deutsche Sammlung von Mikroorganismen (DSM) = 27814) was isolated from a patient with a cervical abscess hospitalized at Hôpital Principal in Dakar, Senegal. *N. massiliensis* is Gram negative, aeroanaerobic, indole negative, nonmotile, and coccobacillus. This bacterium was cultivated as part of the implementation of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) in Hôpital Principal, Dakar, aiming at improving the routine laboratory identification of bacterial strains in Senegal [5].

The current taxonomic classification of prokaryotes relies on a combination of phenotypic and genotypic characteristics [6,7], including 16S rRNA sequence similarity, G+C content and DNA-DNA hybridization. However, these tools suffer from various drawbacks, mainly as a result of their threshold values, which are not applicable to all species or genera [8,9]. With the development of cost-effective high-throughput sequencing techniques, tens of thousands of bacterial genome sequences have been made available in public databases [9]. Recently we developed a strategy, taxonomogenomics, in which genomic and phenotypic characteristics, notably the MALDI-TOF spectrum, are systematically compared to the phylogenetically closest species with standing in nomenclature [8–10].

Here we present a summary classification and a set of features for *Necropsobacter massiliensis* sp. nov. strain FF6^T, together with the description of the complete genomic sequencing and annotation. These characteristics support the circumscription of the species *Necropsobacter massiliensis.*

Organism Information

Classification and features

Since July 2012, the Hôpital Principal in Dakar, Senegal, has been equipped with a MALDI-TOF (Vitek MS RUO;

http://dx.doi.org/10.1016/j.nmni.2015.09.007

bioMérieux, Marcy l'Etoile, France) to improve the microbiology laboratory work flow by enabling rapid bacterial identification. Isolates that are poorly identified using MALDI-TOF are referred to the URMITE laboratory in Marseille, France, for further identification. Strain FF6^T (Table 1) was isolated by cultivation on 5% sheep's blood–enriched Columbia agar (bioMérieux) from the cervical abscess of a 4-year-old Senegalese boy. Strain FF6^T exhibited a 95.17% 16S rRNA sequence identity with *Necropsobacter rosorum* [1], the phylogenetically closest bacterial species with a validly published name (Fig. 1). These values were lower than the 98.7% 16S rRNA gene sequence threshold recommended by Meier-Kolthoff *et al.* [11] to delineate a new species within phylum *Firmicutes* without carrying out DNA-DNA hybridization.

Different growth temperatures (25°C, 30°C, 37°C, 45°C and 56°C) were tested. Growth was obtained between 37°C and 45°C, with the optimal growth temperature being 37°C. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (bioMérieux), and under aerobic conditions with or without 5% CO_2. Optimal growth was observed between 37°C and 45°C under aerobic and microaerophilic conditions. Colonies were 1 mm in diameter, grey and nonhaemolytic on 5% sheep's blood–enriched Columbia agar (bioMérieux). *Necropsobacter massiliensis* is Gram negative, coccobacillus, not motile, and unable to form spores (Fig. 2). Under electron microscopy, cells had a mean length of 1.5 µm (range, 0.9–2.1 µm) and a mean diameter of 0.4 µm (range, 0.2–0.6 µm) (Fig. 3).

Strain FF6^T was oxidase positive and catalase negative. Using an API ZYM strip (bioMérieux), positive reactions were observed for alkaline phosphatase, esterase, leucine arylamidase, phosphatase acid, α-glucosidase and naphthol-AS-BI-phosphohydrolase. Negative reactions were noted for α-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, N-acetyl-β-glucosaminidase, lipase, α-chrymotrypsin and cystine arylamidase. Using API 50CH, positive reactions were observed for glycerol, ribose, D-xylose, D-mannose, D-glucose, inositol, N-acetyl glucosamine, D-fructose, D-maltose D-melibiose, D-trehalose, D-saccharose, D-raffinose, starch, potassium 5-ketogluconate, alkaline phosphatase, esterase, leucine arylamidase, phosphatase acid, α-glucosidase and naphthol-AS-BI-phosphohydrolase. Negative reactions were observed for D-mannitol, D-sorbitol, L-xylose, D-adonitol, methyl β-D-xylopyranose, D-melezitose, inulin, α-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, N-acetyl-β-glucosaminidase, lipase, α-chrymotrypsin and cystine arylamidase. *Necropsobacter massiliensis* strain FF6^T is susceptible to amoxicillin, amoxicillin/clavulanic acid, ceftriaxone, gentamicin, nitrofurantoin, trimethoprim/sulfamethoxazole, rifampicin and ciprofloxacin but resistant to erythromycin, doxycycline and vancomycin. Five species validly published names in the *Pasteurellaceae* family were selected to make a phenotypic comparison with *Necropsobacter massiliensis* (Table 2).

TABLE 1. Classification and general features of *Necropsobacter massiliensis* strain FF6^T

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain: *Bacteria*	TAS [28]
		Phylum: *Proteobacteria*	TAS [29]
		Class: *Gammaproteobacteria*	TAS [30]
		Order: *Pasteurellales*	TAS [31]
		Family: *Pasteurellaceae*	TAS [31,32]
		Genus: *Necropsobacter*	TAS [1]
		Species: *Necropsobacter massiliensis*	IDA
		(Type) strain: FF6^T	IDA
	Gram stain	Negative	IDA
	Cell shape	Rods	IDA
	Motility	None motile	IDA
	Sporulation	Non–spore forming	NAS
	Temperature range	37–45°C	IDA
	Optimum temperature	37°C	IDA
	pH range; optimum	6.2–7.6; 7	
	Carbon source	Unknown	
MIGS-6	Habitat	Human blood	IDA
MIGS-6.3	Salinity	Unknown	
MIGS-22	Oxygen requirement	Aerobic	IDA
MIGS-15	Biotic relationship	Free living	IDA
MIGS-14	Pathogenicity	Unknown	
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection	April 2013	IDA
MIGS-4.1	Latitude	14.6937000	IDA
MIGS-4.1	Longitude	−17.4440600	IDA
MIGS-4.4	Altitude	12 m above sea level	IDA

[a]IDA, inferred from direct assay; TAS, traceable author statement (i.e. a direct report exists in the literature); NAS, nontraceable author statement (i.e. not directly observed for the living, isolated sample but based on a generally accepted property for the species or on anecdotal evidence). These evidence codes are from the Gene Ontology project (http://www.geneontology.org/GO.evidence.shtml) [33]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

Extended features descriptions

MALDI-TOF protein analysis was carried out as previously described [12,13] using a Microflex LT (Bruker Daltonics, Leipzig, Germany). Twelve individual colonies were deposited on a MTP 384 MALDI-TOF target plate (Bruker). A total of 2 µL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 2.5% trifluoroacetic acid were distributed on each smear and air dried for 5 minutes at room temperature. The 12 individual spectra from strain FF6^T were imported into MALDI BioTyper software 2.0 (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6252 bacterial spectra. The scores previously established by Bruker Daltonics allowing (or not) validating the identification of species compared to the database of the instrument were applied. Briefly, a score $\geq$2.000 with a species with a validly published

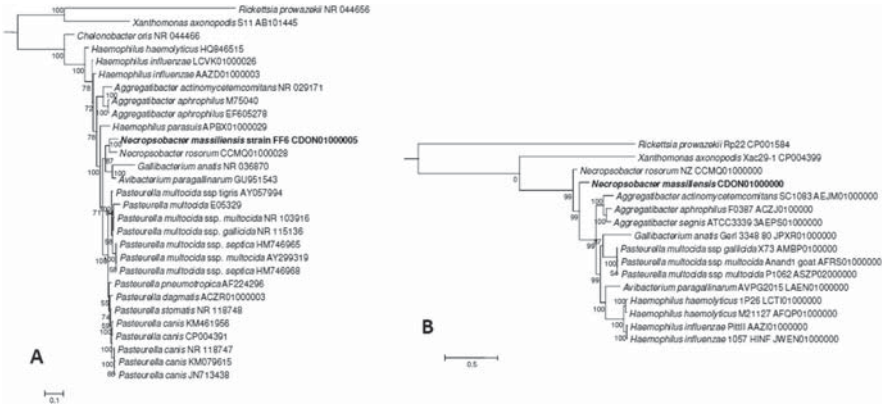


FIG. 1. Phylogenetic trees highlighting position of *Necropsobacter massiliensis* sp. nov. strain FF6 relative to *Pasteurellaceae* type strains. Sequences of 16S rRNA (rrs) gene (A) and concatenated *groEL* and *rpoB* genes (B) were aligned by CLUSTALW, and phylogenetic inferences were obtained from Bayesian phylogenetic analysis. GTR+Ā substitution model was used for rrs-based tree (A) and GTR+Ā, SYM+Ā and GTR+Ā for first, second and third codon position, respectively, for *groEL*/*rpoB* tree (B). GenBank accession numbers of sequences, genomes or shotgun contigs from which gene sequences were extracted are indicated at end. Numbers at nodes are bootstrap values obtained by repeating analysis 100 times to generate majority consensus tree. There were total 1397 (A) and 5814 (1641 for *groEL* and 4173 for *rpoB*) (B) positions in final data set. Scale bar = 10% nucleotide sequence divergence.

name provided allows the identification at the species level; a score of ≥1.700 to <2.000 allows the identification at the genus level; and a score of <1.700 does not allow any identification. Thus, scores ranging from 1.2 to 1.3 were obtained, suggesting that this isolate was not a member of any known species. The reference mass spectrum from strain FF6^T was incremented in our database (Fig. 4). Finally, the gel view showed the spectral differences with other members of the family *Pasteurellaceae* (Fig. 5).

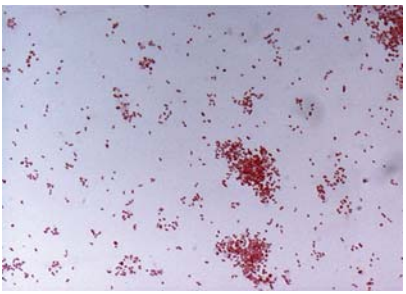


FIG. 2. Gram staining of *Necropsobacter massiliensis* strain FF6^T.

Genome Sequencing Information

Genome project history

The organism was selected for sequencing on the basis of its 16S rRNA similarity, phylogenetic position and phenotypic differences with other members of the family *Pasteurellaceae*, and is part of a study aiming at using MALDI-TOF for the routine identification of bacterial isolates in Hôpital Principal in Dakar [1]. It is the second genome of a *Necropsobacter* species and the first genome of *Necropsobacter massiliensis* sp. nov. A summary of the project information is shown in Table 3. The GenBank accession number is CDON00000000 and consists of 101 contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [14]; associated MIGS records are also summarized in Supplementary Table S1.

Growth conditions and genomic DNA preparation

Necropsobacter massiliensis strain FF6^T (= CSUR P3511 = DSM 27814) was grown aerobically on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C. Bacteria grown on four petri dishes were resuspended in 5 × 100 µL of TE buffer; 150 µL of this suspension was diluted in 350 µL TE buffer

FIG. 3. Transmission electron microscopy of *Necropsobacter massiliensis strain* FF6^T strain. Cells are observed on Tecnai G20 transmission electron microscope operated at 200 keV. Scale bar = 500 µm.

10 × , 25 µL proteinase K and 50 µL sodium dodecyl sulfate for lysis treatment. This preparation was incubated overnight at 56°C. Extracted DNA was purified using three successive phenol–chloroform extractions and ethanol precipitation at −20°C of minimum 2 hours each. After centrifugation, the DNA was suspended in 65 µL EB buffer. The genomic DNA concentration was measured at 30.06 ng/µL using the Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA).

Genome sequencing and assembly

Genomic DNA of *Necropsobacter massiliensis* FF6^T was sequenced on the MiSeq sequencer (Illumina, San Diego, CA, USA) with the Paired-end and Mate-pair strategies. The paired-end and the mate-pair strategies were barcoded in order to be mixed respectively with ten other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina).

Genomic DNA was diluted to 1 ng/µL to prepare the paired-end library. The "tagmentation" step fragmented and tagged the DNA with an optimal size distribution at 1.5 kb. Then limited-cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter, Fullerton, CA, USA), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single-strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in single 39-hour runs at 2 × 250 bp.

Total information of 3.89 GB was obtained from a 416 k/mm^2 cluster density with a cluster passing quality control filters of 95.4% (7 899 000 clusters). Within this run, the index representation for *Necropsobacter massiliensis* was determined to

TABLE 2. **Differential characteristics of *Necropsobacter massiliensis* strain FF6^T with *Necropsobacter rosorum* [1], *Actinobacillus actinomycetemcomitans* [34,35], *Haemophilus influenzae* [34–36] and *Pasteurella multocida* [34–37]**

Character	*Necropsobacter massiliensis*	*Necropsobacter rosorum*	*Actinobacillus actinomycetemcomitans*	*Haemophilus influenzae*	*Pasteurella multocida*
Cell diameter (µm)	0.4	NA	0.5		
Gram stain	−	−	−	−	−
Motility	−	−	−	−	−
Endospore formation	−	NA	−	NA	NA
Production of					
Alkaline phosphatase	+	NA	+	+	Variable
Acid phosphatase	+	NA	+	NA	Variable
Catalase	−	+	+	+	+
Oxidase	+	+	−	+	+
β-Haemolysis	−	−	−	−	−
Urease	−	−	−	+	−
Indole	−	−	NA	+	+
Nitrate reductase	+	+	+	+	+
α-Galactosidase	−	+	NA	−	NA
β-Galactosidase					
α-glucosidase (PNPG)	+	+	NA	−	NA
β-Glucosidase	−	−	NA	−	NA
Esterase	+	NA	Variable	−	Variable
Esterase lipase	−	NA	Variable	NA	Variable
N-acetyl-β-glucosaminidase	−	NA	NA	NA	NA
Utilization of					
D-Fructose	+	−	+	NA	+
D-Mannose	+	+	+	NA	+
D-Xylose	+	+	Variable	+	+
D-Glucose	+	+	+	+	NA
Habitat	Human	Guinea pig	Human	Human	Human and animal

NA, data not available.

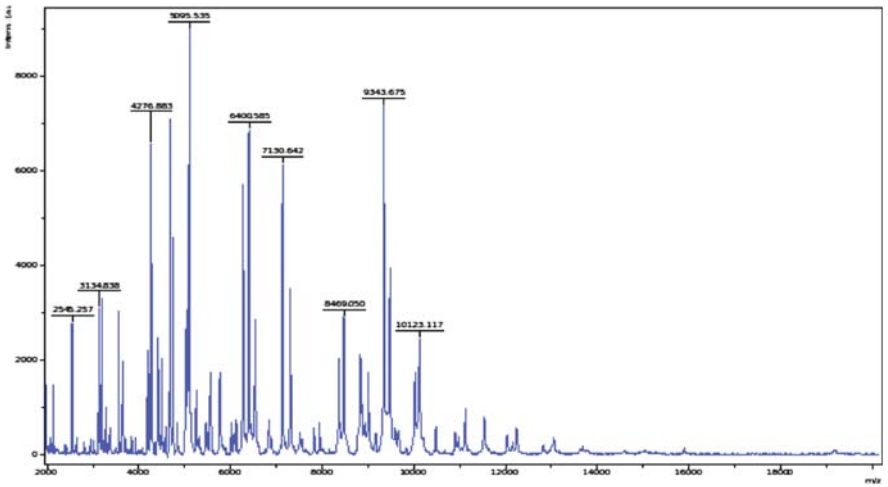


FIG. 4. Reference mass spectrum from *Necropsobacter massiliensis* strain FF6^T. Spectra from 12 individual colonies were compared and reference spectrum was generated.

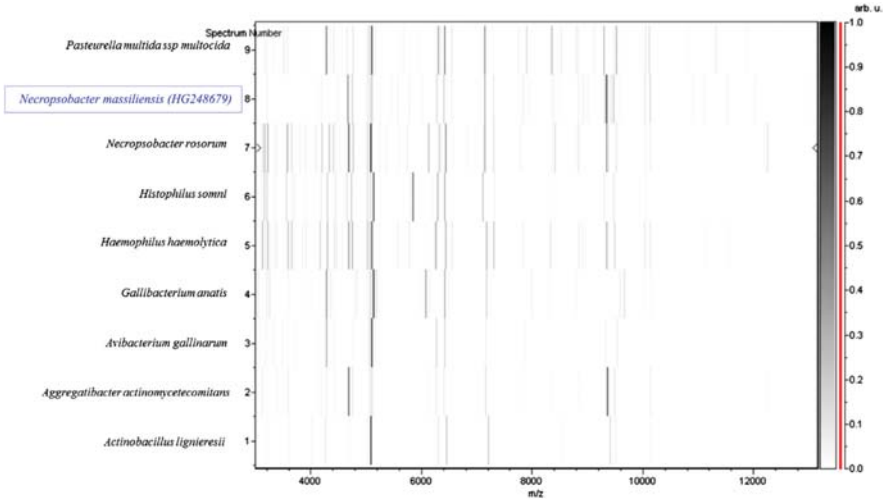


FIG. 5. Gel view comparing *Necropsobacter massiliensis* strain FF6^T to members of family *Pasteurellaceae*. Gel view displays raw spectra of all loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value; left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Colour bar and right y-axis indicating relation between colour peak is displayed; peak intensity indicated arbitrary units. Displayed species are indicated at left.

TABLE 3. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Paired end and mate pair 9 kb library
MIGS-29	Sequencing platforms	MiSeq
MIGS-31.2	Fold coverage	141×
MIGS-30	Assemblers	CLC 7
MIGS-32	Gene calling method	Prodigal
	Locus Tag	Not indicated
	GenBank ID	CDON00000000
	GenBank date of release	26 March 2015
	GOLD ID	Gp0102103
	BIOPROJECT	PRJEB4626
	Source material identifier	DSM 27814
	Project relevance	MALDI-TOF implementation in Dakar

MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

be 7.02% and to present 529 002 reads filtered according to the read qualities.

The mate-pair library was prepared with 1 µg of genomic DNA using the Nextera mate-pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1 kb up to 6.6 kb, with an optimal size at 4.5 kb. No size selection was performed, and 368 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments on the Covaris S2 device in microtubes (Covaris, Woburn, MA, USA).The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent) with an optimal peak at 672 bp. The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 10 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-hour run at 2 × 250 bp. *Necropsobacter massiliensis* strain FF6^T was determined to be 6.86%. The 639 775 reads were filtered according to the read qualities.

Genome annotation

Open reading frame (ORF) prediction was carried out using Prodigal [15] with default parameters. We removed the predicted ORFs if they spanned a sequencing gap region. Functional assessment of protein sequences was performed by comparing them with sequences in the GenBank [16] and the Clusters of Orthologous Groups (COGs) databases using BLASTP. tRNAs, rRNAs, signal peptides and transmembrane helices were identified using tRNAscan-SE 1.21 [17], RNAmmer [18], SignalP [19] and TMHMM [20], respectively. Artemis [21] was used for data management, whereas DNA Plotter [22] was used for visualization of genomic features. In-house perl and bash scripts were used to automate these routine tasks. ORFans were sequences which have no homology in a given database, *i.e.* in nonredundant (nr) or identified if their BLASTP E value was lower than 1e-03 for alignment lengths greater than 80 amino acids. PHAST was used to identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids [23].

To estimate the nucleotide sequence similarity at the genome level between *Necropsobacter massiliensis* and another ten members of the *Pasteurellaceae* family, we determined the average genomic identity of orthologous gene sequences (AGIOS) parameter as follows: orthologous proteins were detected using the Proteinortho software [24] (with the following parameters: E-value 1e-5, 30% percentage identity, 50% coverage and 50% algebraic connectivity) and genomes compared two by two. After fetching the corresponding nucleotide sequences of orthologous proteins for each pair of genomes, we determined the mean percentage of nucleotide sequence identity using the Needleman-Wunsch global alignment algorithm. Genomes from the genus *Necropsobacter* and closely related genera were used for the calculation of AGIOS values. The script created to calculate AGIOS values was named MAGi (Marseille Average genomic identity) and is written in perl and bioperl modules. GGDC analysis was also performed using the GGDC Web server (http://ggdc.dsmz.de) as previously reported [25,26].

Here, we compared the genome sequences of *Necropsobacter massiliensis* strain FF6^T (GenBank accession number CDON00000000) with those of *N. rosorum* strain P709T (CCMQ00000000), *Pasteurella multocida* subsp. *multocida* strain Pm70 (AE004439), *Haemophilus influenzae* strain Rd KW20 (L42023), *Haemophilus ducreyi* strain 35000HP (AE017143), *Histophilus somnus* strain 129PT (CP000436), *Haemophilus parasuis* strain SH0165 (CP001321), *Haemophilus parainfluenzae* strain T3T1 (FQ312002) and *Aggregatibacter aphrophilus* strain NJ8700 (CP001607).

Genome properties

The genome of *Necropsobacter massiliensis* strain FF6^T is 2 493 927 bp long with a 46.2% G+C content (Fig. 6). Of the 2363 predicted genes, 2309 were protein coding genes and 54 were RNA genes including 1 complete rRNA operon. A total of 1838 genes (77.7%) were assigned a putative function. A total of 210 were identified as ORFans (9.09%). The remaining genes were annotated as hypothetical proteins. The properties and the statistics of the genome are presented in Table 4. The distribution of genes into COGs functional categories is summarized in Table 5.

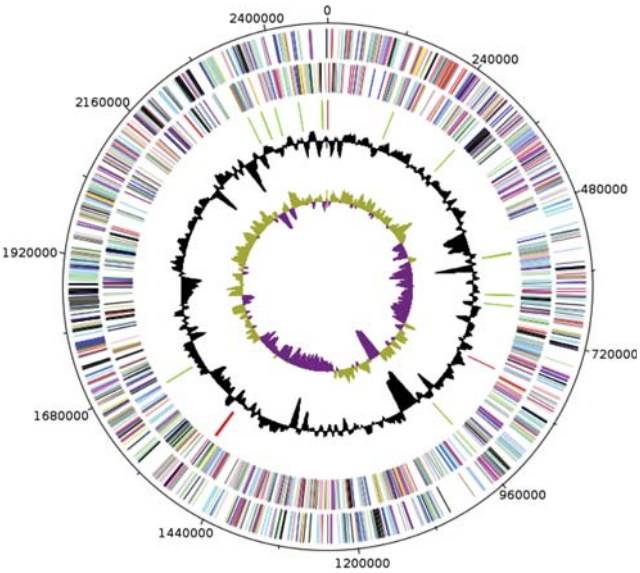


FIG. 6. Graphical circular map of *Necropsobacter massiliensis* strain FF6^T chromosome. From outside in, outer two circles shows open reading frames oriented in forward (coloured by COGs categories) and reverse (coloured by COGs categories) direction, respectively. Third circle marks indicate rRNA gene operon (green) and tRNA genes (red). Fourth circle shows G+C% content plot. Innermost circle shows GC skew, with purple indicating negative values and olive positive values.

Insights From Genome Sequence

Extended insights

The draft genome of *Necropsobacter massiliensis* (2.49 Mb) has a lower size than that of *N. rosorum* (2.52 Mb) but a larger size than those of *P. multocida* (2.25 Mb), *H. influenzae* (1.83 Mb), *H. ducreyi* (1.69 Mb), *H. somnus* (2.00 Mb), *H. parasuis* (2.26 Mb), *H. aphrophilus* (2.31 Mb) and *H. parainfluenzae* (2.08 Mb). The G+C content of *Necropsobacter massiliensis* (46.2%) was lower than that of *N. rosorum* (48.9%) but higher than those of *P. multocida* (40.40%), *H. influenzae* (38.15%), *H. ducreyi* (38.22%), *H. somnus* (37.20%), *H. parasuis* (39.99%), *H. aphrophilus* (42.23%) and *H. parainfluenzae* (39.57%). Because it has been suggested in the literature that the G+C content deviation is at most 1% within species, these data are an additional argument for the creation of a new taxon [27].

TABLE 4. **Genome information**

Attribute	Value	% of total
Genome size (bp)	2 493 927	
DNA coding (bp)	2 230 337	89.4
DNA G+C (bp)	1 151 339	46.2
DNA scaffolds	43	
Total genes	2363	100
Protein coding genes	2309	97.7
RNA genes	54	
Pseudo genes	Not indicated	
Genes in internal clusters	130	5.63
Genes with function prediction	1838	77.7
Genes assigned to COGs	2035	88.1
Genes with Pfam domains	75	3.24
Genes with signal peptides	210	9.09
Genes with transmembrane helices	561	24.3
CRISPR repeats	3	

COGs, Clusters of Orthologous Groups.

The protein-coding genes of *Necropsobacter massiliensis* is larger than those of *P. multocida, H. influenzae, H. ducreyi, H. somnus, H. parasuis, H. aphrophilus* and *H. parainfluenzae* (2012, 1603, 1717, 1791, 2021, 2218 and 1975, respectively) but smaller than that of *N. rosorum* (2311). However, the distribution of genes into categories was similar in all compared genomes. In addition, *Necropsobacter massiliensis* shared 2012, 1603, 1717, 1791, 2021, 1975, 2301 and 2218 orthologous genes with *P. multocida, H. influenzae, H. ducreyi, H. somnus, H. parasuis, H. aphrophilus* and *H. parainfluenzae,* respectively. Among species with standing in nomenclature, AGIOS values ranged from 66.32 between *N. rosorum* and *H. ducreyi* to 98.71% between *P. multocida* and *H. parainfluenzae* (Table 6). When

TABLE 5. Number of genes associated with general COGs functional categories[a]

Code	Value	Percentage	Description
J	152	6.58	Translation, ribosomal structure and biogenesis
A	1	0.04	RNA processing and modification
K	100	4.33	Transcription
L	127	5.50	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	26	1.13	Cell cycle control, cell division, chromosome partitioning
V	21	0.91	Defense mechanisms
T	29	1.26	Signal transduction mechanisms
M	117	5.07	Cell wall/membrane biogenesis
N	0	0.00	Cell motility
U	37	1.60	Intracellular trafficking and secretion
O	86	3.72	Posttranslational modification, protein turnover, chaperones
C	113	4.89	Energy production and conversion
G	182	7.88	Carbohydrate transport and metabolism
E	152	6.58	Amino acid transport and metabolism
F	55	2.38	Nucleotide transport and metabolism
H	86	3.72	Coenzyme transport and metabolism
I	44	1.91	Lipid transport and metabolism
P	112	4.85	Inorganic ion transport and metabolism
Q	8	0.35	Secondary metabolites biosynthesis, transport and catabolism
R	198	8.58	General function prediction only
S	172	7.45	Function unknown
—	197	8.53	Not in COGs

Total is based on total number of protein coding genes in annotated genome.
COGs, Clusters of Orthologous Groups.

compared to other species, *Necropsobacter massiliensis* exhibited AGIOS values ranging from 67.15 with *H. ducreyi* to 84.44 with *N. rosorum*. We obtained similar results using the GGDC software, as dDDH values ranged from 0.201 to 0.281 between studied species and were 0.275 between *N. rosorum*. These values confirm the status of *Necropsobacter massiliensis* as a new species.

Conclusions

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Necropsobacter massiliensis* sp. nov. that contains strain FF6^T. The strain was isolated from a cervical abscess of a 4-year-old Senegalese boy.

Taxonomic and nomenclatural proposals: description of *Necropsobacter massiliensis* strain FF6^T sp. nov.

Necropsobacter massiliensis (mas•il•ien'sis, L. gen., fem. n. *massiliensis*, of Massilia, the Latin name of Marseille, where this strain was characterized). On 5% sheep's blood–enriched Columbia agar (BioMérieux), colonies were 1 mm in diameter and grey. Cells are Gram negative and not motile, with a mean diameter of 0.4 µm (range, 0.2–0.6 µm) and a mean length of 1.5 µm (range, 0.9–2.1 µm). Catalase test was negative and oxidase test was positive. Positive reactions were observed for glycerol, ribose, D-xylose, D-mannose, D-glucose, inositol, N-acetyl glucosamine, D-fructose, D-maltose D-melibiose, D-trehalose, D-saccharose, D-raffinose, starch, potassium 5-ketogluconate, alkaline phosphatase, esterase, leucine arylamidase, phosphatase acid, α-glucosidase, naphthol-AS-BI-phosphohydrolase and α-glucosidase. Negative reactions were observed for D-mannitol, D-sorbitol, L-xylose, D-adonitol, methyl β-D-xylopyranoside, D-melezitose, inulin, α-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, N-acetyl-β-glucosaminidase, lipase, α-chrymotrypsin, cystine arylamidase, α-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, N-acetyl-β-glucosaminidase, lipase, α-chrymotrypsin and cystine arylamidase.

Necropsobacter massiliensis strain FF6^T is susceptible to amoxicillin, amoxicillin/clavulanic acid, ceftriaxone, gentamicin, nitrofurantoin, rifampicin, trimethoprim/sulfamethoxazole and ciprofloxacin but resistant to erythromycin, doxycycline and vancomycin. The G+C content of the genome is 46.2%. The 16S rRNA and genome sequences of *N. massiliensis* strain FF6^T (= CSUR P3511 = DSM 27814) are deposited in GenBank under accession numbers HG428679 and CDON00000000, respectively. The type strain, FF6^T, was isolated from a cervical abscess of a 4-year-old Senegalese boy hospitalized in Hôpital Principal in Dakar, Senegal.

TABLE 6. Number of orthologous proteins shared between genomes (upper right) and AGIOS values obtained (lower left)

	N. massiliensis	*P. multocida.* subsp. *multocida*	*H. influenzae*	*H. ducreyi*	*H. somnus*	*H. parasuis*	*H. parainfluenzae*	*N. rosorum*	*A. aphrophilus*
N. massiliensis	2311[a]	71.37	70.61	67.15	71.73	68.01	71.29	84.44	72.12
P. multocida. subsp. *multocida*	1508	2012[a]	72.86	69.26	73.13	69.78	98.71	70.73	72.45
H. influenzae	1305	1271	1603[a]	69.89	72.96	70.15	72.81	69.71	73.69
H. ducreyi	1137	1107	1013	1717[a]	69.74	73.04	69.26	66.32	68.61
H. somnus	1306	1261	1124	1010	1791[a]	69.88	73.12	71.00	71.88
H. parasuis	1335	1290	1165	1068	1140	2021[a]	69.63	67.23	69.35
H. parainfluenzae	1523	1869	1275	1119	1270	1307	1975[a]	70.68	72.40
N. rosorum	1730	1514	1323	1111	1236	1335	1528	2301[a]	71.90
A. aphrophilus	1514	1407	1239	1065	1183	1204	1412	1463	2218[a]

Necropsobacter massiliensis FF6^T; *Necropsobacter rosorum*; *Pasteurella multocida* subsp. *multocida* Pm70; *Haemophilus influenzae* Rd KW20; *Haemophilus ducreyi* 35000HP; *Haemophilus somnus* 129PT; *Haemophilus parasuis* SH0165; HPI^g, *Haemophilus parainfluenzae* T3T1; *Aggregatibacter aphrophilus* NJ8700.
AGIOS, average genomic identity of orthologous gene sequences.
[a]Numbers of proteins per genome.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process. This study was funded by the Mediterranée-Infection Foundation. We thank J. Rathored for taking part in this work.

Conflict of Interest

None declared.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.nmni.2015.09.007.

References

[1] Christensen H, Korczak BM, Bojesen AM, Kuhnert P, Frederiksen W, Bisgaard M. Classification of organisms previously reported as the SP and Stewart-Letscher groups, with descriptions of *Necropsobacter* gen. nov. and of *Necropsobacter rosorum* sp. nov. for organisms of the SP group. Int J Syst Evol Microbiol 2011;61:1829–36.
[2] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 2014;42:D613–6.
[3] Tan TY, Ng LS, Eng LC. Five clinical cases of *Necropsobacter rosorum* bacteremia. J Clin Microbiol 2013;51:1337–40.
[4] Padmanabhan R, Robert C, Fenollar F, Raoult D, Fournier PE. Draft genome sequence of *Necropsobacter rosorum* strain P709T. Genome Announc 2014;2.
[5] Fall B, Lo CI, Samb-Ba B, et al. The ongoing revolution of MALDI-TOF mass spectrometry for microbiology reaches tropical Africa. Am J Trop Med Hyg 2015;92:641–7.
[6] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[7] Ramasamy D, Mishra AK, Lagier JC, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[8] Sentausa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect 2013;19:790–5.
[9] Reddy TBK, Thomas A, Stamatis D, et al. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification. Nucleic Acids Res 2014;43:1099–106.
[10] Lo CI, Mishra AK, Padhmanabhan R, et al. Non-contiguous finished genome sequence and description of *Clostridium dakarense* sp. nov. Stand Genomic Sci 2013;30:14–27.
[11] Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 2013;195:413–8.
[12] Seng P, Drancourt M, Gouriet F, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.
[13] Seng P, Rolain JM, Fournier PE, La Scola B, Drancourt M, Raoult D. MALDI-TOF–mass spectrometry applications in clinical microbiology. Future Microbiol 2010;5:1733–54.
[14] Field D, Garrity G, Gray T, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26:541–7.
[15] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.
[16] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:48–53.
[17] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[18] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucl Acids Res 2007;35:3100–8.
[19] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[20] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.
[21] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[22] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[23] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.
[24] Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.
[25] Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci 2010;2:142–8.
[26] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence–based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013;14:60.
[27] Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 2014;64:352–6.
[28] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eukarya. Proc Natl Acad Sci U S A 1990;87:4576–9.
[29] Garrity GM, Bell JA, Lilburn T, Phylum XIV. *Proteobacteria* phyl. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. Bergeys's manual of systematic bacteriology (2nd ed., vol. 2, pt. B). New York: Springer; 2005. p. 1.
[30] Williams KP, Kelly DP. Proposal for a new class within the phylum *Proteobacteria, Acidithiobacillia* classis nov., with the type order *Acidithiobacillales*, and emended description of the class *Gammaproteobacteria*. Int J Syst Evol Microbiol 2013;63:2901–6.
[31] Christensen H, Kuhnert P, Busse HJ, Frederiksen WC, Bisgaard M. Proposed minimal standards for the description of genera, species and subspecies of the *Pasteurellaceae*. Int J Syst Evol Microbiol 2007;57:166–78.
[32] Pohl S. Family *Pasteurellaceae* fam. nov. in validation of the publication of new names and new combinations previously effectively published outside the IJSB, list no. 7. Int J Syst Bacteriol 1981;31:382–3.
[33] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
[34] Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol 1980;30:225–420.

[35] Nørskov-Lauritsen N, Kilian M. Reclassification of *Actinobacillus actinomycetemcomitans, Haemophilus aphrophilus, Haemophilus paraphrophilus and Haemophilus segnis* as *Aggregatibacter actinomycetemcomitans* gen. nov., comb. nov., *Aggregatibacter aphrophilus* comb. nov. and *Aggregatibacter segnis* comb. nov., and emended description of *Aggregatibacter aphrophilus* to include V factor–dependent and V factor–independent isolates. Int J Syst Evol Microbiol 2006;56:2135–46.

[36] Winslow CEA, Broadhurst J, Buchanan RE, Krumwiede C, Rogers LA, Smith GH. The families and genera of the bacteria preliminary report of the Committee of the Society of American Bacteriologists on characterization and classification of bacterial types. J Bacteriol 1917;2: 505–66.

[37] Mutters R, Ihm P, Pohl S, Frederiksen W, Mannheim W. Reclassification of the genus *Pasteurella* Trevisan 1887 on the basis of deoxyribonucleic acid homology, with proposals for the new species *Pasteurella dagmatis, Pasteurella canis, Pasteurella stomatis, Pasteurella anatis*, and *Pasteurella langaa*. Int J Syst Bacteriol 1985;35: 309–22.

Article 7:

« Non-contiguous finished genome sequence and description of Necropsobacter massiliensis sp. nov. »

Cheikh Ibrahima Lo[1], Bécaye Fall[2], Roshan Padhmanabhan[1], Bissoume Samb-Ba[2], Didier Raoult[1], Pierre-Edouard Fournier[1] and Florence Fenollar[1]*

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France Tel: +33 491 385 517
[2]Hôpital Principal, Dakar, Senegal

Publié dans :

New Microbes and New Infections

November 2015 Volume 8, Pages 41-50

http://dx.doi.org/10.1016/j.nmni.2015.09.007

Résumé de l'article 7: ***« Non-contiguous finished genome sequence and description of Necropsobacter massiliensis sp. nov. »***

La souche FF6^T a été isolée d'un abcès cervical d'un garçon Sénégalais de 4 ans à l'Hôpital Principal de Dakar au Sénégal. L'analyse par spectrométrie de masse MALDI-TOF n'a pas permis d'identification. La souche FF6^T a une similarité de séquence de l'ARN 16S ribosomique de 95,17% avec *Necropsobacter rosorum*, l'espèce officielle phylogénétiquement la plus proche. Une étude polyphasique basée sur des analyses phénotypiques et génomiques a été réalisée. La souche FF6^T est une bactérie à Gram négatif, aéro-anaérobie, en forme de coccobacille. La bactérie est immobile. Les cellules ont un diamètre moyen de 0,4 µm (entre 0,2-0,6 µm) et une longueur moyenne de 1,5 µm (entre 0,9-2,1 µm). Sur gélose Columbia enrichi de 5% de sang de mouton, les colonies ont 1 mm de diamètre et sont grises. La catalase est négative et l'oxydase positive. En utilisant la galerie API 50CH et ZYM, des réactions positives sont aussi observées pour le glycérol, le ribose, le D-xylose, le D-mannose, le D-glucose, l'inositol, la N-Acetyl glucosamine, le D-fructose, le D-maltose, le D-mélibiose, le D-tréhalose, le D-saccharose, le D-raffinose, l'amidon, potassium 5-cétoGluconate, l'acide phosphatase, naphthol-AS-BI-phosphohydrolase, l'alcaline phosphatase, l'estérase, la leucine arylamidase et l'α-glucosidase.

La souche FF6^T est sensible *in vitro* à l'amoxicilline, l'amoxicilline/acide clavulanique, la ceftriaxone, la gentamicine, le nitrofurantoine, la rifampicine, le triméthoprime/sulfaméthoxazole et la ciprofloxacine mais résistante à l'érythromycine, la doxycycline et la

vancomycine. Le génome est de 2 493 927 paires de base comprenant 2 309 gènes codant pour des protéines et 53 gènes ARN (y compris 1 opéron ARNr). Le contenu en G+C est de 46,2%.

Sur la base de ces données, nous proposons formellement la création de *Necropsobacter massiliensis* sp. nov.

Description de *Necropsobacter massiliensis sp. nov.*

Necropsobacter massiliensis (mas.il.ien'sis L. gen. Fem. n. massiliensis, de Massilia, le nom Latin de Marseille, ville où cette souche a été caractérisée). Les séquences de l'ARN 16S ribosomique et du génome de *N. massiliensis* strain FF6^T (=CSUR P3511 =DSM 27814) ont été déposées dans GenBank avec les numéros d'accession respectifs HG428679 et CDON00000000. La souche type FF6^T a été isolée d'un abcès cervical d'un garçon de 4 ans hospitalisé à l'Hôpital Principal de Dakar, Sénégal.

Noncontiguous finished genome sequence and description of *Necropsobacter massiliensis* sp. nov.

C. I. Lo[1,3], R. Padhamanabhan[1], B. Fall[2], B. Sambe-Ba[2], O. Mediannikov[1,3], T.-T. Nguyen[1], E. Prudent[1], N. Faye[4], B. Wade[2], D. Raoult[1,3,5], P.-E. Fournier[1] and F. Fenollar[1,3]

1) *Faculté de médecine, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Aix-Marseille Université, Marseille, France,* 2) *Hôpital Principal,* 3) *Campus International UCAD-IRD,* 4) *Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Dakar, Senegal and* 5) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Strain FF6^T was isolated from the cervical abscess of a 4-year-old Senegalese boy, in Dakar, Senegal. MALDI-TOF MS did not provide any identification. This strain exhibited a 95.17% 16S rRNA sequence identity with *Necropsobacter rosorum*. Using a polyphasic study including phenotypic and genomic analyses, strain FF6^T was an aero-anaerobic Gram-negative cocobacillus, oxidase positive, and exhibited a genome of 2,493,927 bp (1 chromosome but no plasmid) with a G+C content of 46.2% that coded 2,309 protein-coding and 53 RNA genes. On the basis of these data, we propose the creation of *Necropsobacter massiliensis* sp. nov.

Keywords: Culturomics, genome, *Necropsobacter massiliensis*, Senegal, taxono-genomics

Original Submission: 15 June 2015; **Revised Submission:** 4 September 2015; **Accepted:** 7 September 2015

Article published online: 16 September 2015

Corresponding author: F. Fenollar, Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France

E-mail: florence.fenollar@univ-amu.fr

Introduction

The genus *Necropsobacter* (Christensen *et al.* 2011) was first described in 2011 [1]. At this time, there is only one species with a validly published name [2]. In 2013, five clinical cases of bacteraemia associated with *Necropsobacter rosorum* were reported [3]. Members of the genus *Necropsobacter* were previously associated with the SP group that comprised mainly strains isolated from rabbits, rodents and humans [3]. Because *Necropsobacter rosorum* was the only described species in this genus with no genome available, we first sequenced its genome for genomic comparison [4]. *Necropsobacter massiliensis* strain FF6^T (= Collection de souches de l'Unité des Rickettsies (CSUR) P3511 = Deutsche Sammlung von Mikroorganismen (DSM) = 27814) was isolated from a patient with a cervical abscess hospitalized at Hôpital Principal in Dakar, Senegal. *N. massiliensis* is Gram negative, aeroanaerobic, indole negative, nonmotile, and coccobacillus. This bacterium was cultivated as part of the implementation of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) in Hôpital Principal, Dakar, aiming at improving the routine laboratory identification of bacterial strains in Senegal [5].

The current taxonomic classification of prokaryotes relies on a combination of phenotypic and genotypic characteristics [6,7], including 16S rRNA sequence similarity, G+C content and DNA-DNA hybridization. However, these tools suffer from various drawbacks, mainly as a result of their threshold values, which are not applicable to all species or genera [8,9]. With the development of cost-effective high-throughput sequencing techniques, tens of thousands of bacterial genome sequences have been made available in public databases [9]. Recently we developed a strategy, taxonomogenomics, in which genomic and phenotypic characteristics, notably the MALDI-TOF spectrum, are systematically compared to the phylogenetically closest species with standing in nomenclature [8–10].

Here we present a summary classification and a set of features for *Necropsobacter massiliensis* sp. nov. strain FF6^T, together with the description of the complete genomic sequencing and annotation. These characteristics support the circumscription of the species *Necropsobacter massiliensis.*

Organism Information

Classification and features

Since July 2012, the Hôpital Principal in Dakar, Senegal, has been equipped with a MALDI-TOF (Vitek MS RUO;

http://dx.doi.org/10.1016/j.nmni.2015.09.007

bioMérieux, Marcy l'Etoile, France) to improve the microbiology laboratory work flow by enabling rapid bacterial identification. Isolates that are poorly identified using MALDI-TOF are referred to the URMITE laboratory in Marseille, France, for further identification. Strain FF6^T (Table 1) was isolated by cultivation on 5% sheep's blood–enriched Columbia agar (bioMérieux) from the cervical abscess of a 4-year-old Senegalese boy. Strain FF6^T exhibited a 95.17% 16S rRNA sequence identity with *Necropsobacter rosorum* [1], the phylogenetically closest bacterial species with a validly published name (Fig. 1). These values were lower than the 98.7% 16S rRNA gene sequence threshold recommended by Meier-Kolthoff *et al.* [11] to delineate a new species within phylum *Firmicutes* without carrying out DNA-DNA hybridization.

Different growth temperatures (25°C, 30°C, 37°C, 45°C and 56°C) were tested. Growth was obtained between 37°C and 45°C, with the optimal growth temperature being 37°C. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (bioMérieux), and under aerobic conditions with or without 5% CO_2. Optimal growth was observed between 37°C and 45°C under aerobic and microaerophilic conditions. Colonies were 1 mm in diameter, grey and nonhaemolytic on 5% sheep's blood–enriched Columbia agar (bioMérieux). *Necropsobacter massiliensis* is Gram negative, coccobacillus, not motile, and unable to form spores (Fig. 2). Under electron microscopy, cells had a mean length of 1.5 µm (range, 0.9–2.1 µm) and a mean diameter of 0.4 µm (range, 0.2–0.6 µm) (Fig. 3).

Strain FF6^T was oxidase positive and catalase negative. Using an API ZYM strip (bioMérieux), positive reactions were observed for alkaline phosphatase, esterase, leucine arylamidase, phosphatase acid, α-glucosidase and naphthol-AS-BI-phosphohydrolase. Negative reactions were noted for α-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, N-acetyl-β-glucosaminidase, lipase, α-chrymotrypsin and cystine arylamidase. Using API 50CH, positive reactions were observed for glycerol, ribose, D-xylose, D-mannose, D-glucose, inositol, N-acetyl glucosamine, D-fructose, D-maltose D-melibiose, D-trehalose, D-saccharose, D-raffinose, starch, potassium 5-ketogluconate, alkaline phosphatase, esterase, leucine arylamidase, phosphatase acid, α-glucosidase and naphthol-AS-BI-phosphohydrolase. Negative reactions were observed for D-mannitol, D-sorbitol, L-xylose, D-adonitol, methyl β-D-xylopyranose, D-melezitose, inulin, α-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, N-acetyl-β-glucosaminidase, lipase, α-chrymotrypsin and cystine arylamidase. *Necropsobacter massiliensis* strain FF6^T is susceptible to amoxicillin, amoxicillin/clavulanic acid, ceftriaxone, gentamicin, nitrofurantoin, trimethoprim/sulfamethoxazole, rifampicin and ciprofloxacin but resistant to erythromycin, doxycycline and vancomycin. Five species validly published names in the *Pasteurellaceae* family were selected to make a phenotypic comparison with *Necropsobacter massiliensis* (Table 2).

TABLE 1. Classification and general features of *Necropsobacter massiliensis* strain FF6^T

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain: *Bacteria*	TAS [28]
		Phylum: *Proteobacteria*	TAS [29]
		Class: *Gammaproteobacteria*	TAS [30]
		Order: *Pasteurellales*	TAS [31]
		Family: *Pasteurellaceae*	TAS [31,32]
		Genus: *Necropsobacter*	TAS [1]
		Species: *Necropsobacter massiliensis*	IDA
		(Type) strain: FF6^T	IDA
	Gram stain	Negative	IDA
	Cell shape	Rods	IDA
	Motility	None motile	IDA
	Sporulation	Non–spore forming	NAS
	Temperature range	37–45°C	IDA
	Optimum temperature	37°C	IDA
	pH range; optimum	6.2–7.6; 7	
	Carbon source	Unknown	
MIGS-6	Habitat	Human blood	IDA
MIGS-6.3	Salinity	Unknown	
MIGS-22	Oxygen requirement	Aerobic	IDA
MIGS-15	Biotic relationship	Free living	IDA
MIGS-14	Pathogenicity	Unknown	
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection	April 2013	IDA
MIGS-4.1	Latitude	14.6937000	IDA
MIGS-4.1	Longitude	−17.4440600	IDA
MIGS-4.4	Altitude	12 m above sea level	IDA

[a]IDA, inferred from direct assay; TAS, traceable author statement (i.e. a direct report exists in the literature); NAS, nontraceable author statement (i.e. not directly observed for the living, isolated sample but based on a generally accepted property for the species or on anecdotal evidence). These evidence codes are from the Gene Ontology project (http://www.geneontology.org/GO.evidence.shtml) [33]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

Extended features descriptions

MALDI-TOF protein analysis was carried out as previously described [12,13] using a Microflex LT (Bruker Daltonics, Leipzig, Germany). Twelve individual colonies were deposited on a MTP 384 MALDI-TOF target plate (Bruker). A total of 2 µL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 2.5% trifluoroacetic acid were distributed on each smear and air dried for 5 minutes at room temperature. The 12 individual spectra from strain FF6^T were imported into MALDI BioTyper software 2.0 (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6252 bacterial spectra. The scores previously established by Bruker Daltonics allowing (or not) validating the identification of species compared to the database of the instrument were applied. Briefly, a score $\geq$2.000 with a species with a validly published

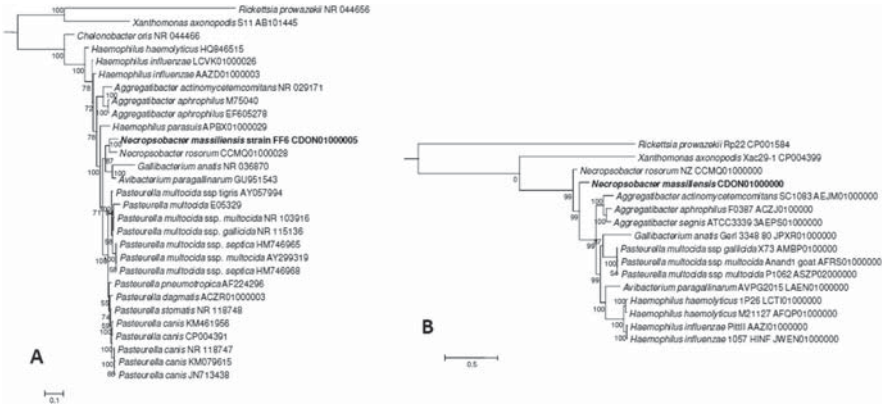


FIG. 1. Phylogenetic trees highlighting position of *Necropsobacter massiliensis* sp. nov. strain FF6 relative to *Pasteurellaceae* type strains. Sequences of 16S rRNA (rrs) gene (A) and concatenated *groEL* and *rpoB* genes (B) were aligned by CLUSTALW, and phylogenetic inferences were obtained from Bayesian phylogenetic analysis. GTR+Ā substitution model was used for rrs-based tree (A) and GTR+Ā, SYM+Ā and GTR+Ā for first, second and third codon position, respectively, for *groEL*/*rpoB* tree (B). GenBank accession numbers of sequences, genomes or shotgun contigs from which gene sequences were extracted are indicated at end. Numbers at nodes are bootstrap values obtained by repeating analysis 100 times to generate majority consensus tree. There were total 1397 (A) and 5814 (1641 for *groEL* and 4173 for *rpoB*) (B) positions in final data set. Scale bar = 10% nucleotide sequence divergence.

name provided allows the identification at the species level; a score of $\geq$1.700 to <2.000 allows the identification at the genus level; and a score of <1.700 does not allow any identification. Thus, scores ranging from 1.2 to 1.3 were obtained, suggesting that this isolate was not a member of any known species. The reference mass spectrum from strain FF6^T was incremented in our database (Fig. 4). Finally, the gel view showed the spectral differences with other members of the family *Pasteurellaceae* (Fig. 5).

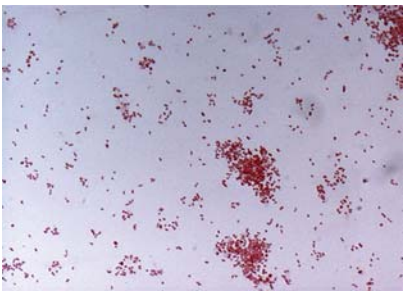


FIG. 2. Gram staining of *Necropsobacter massiliensis* strain FF6^T.

Genome Sequencing Information

Genome project history

The organism was selected for sequencing on the basis of its 16S rRNA similarity, phylogenetic position and phenotypic differences with other members of the family *Pasteurellaceae*, and is part of a study aiming at using MALDI-TOF for the routine identification of bacterial isolates in Hôpital Principal in Dakar [1]. It is the second genome of a *Necropsobacter* species and the first genome of *Necropsobacter massiliensis* sp. nov. A summary of the project information is shown in Table 3. The GenBank accession number is CDON00000000 and consists of 101 contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [14]; associated MIGS records are also summarized in Supplementary Table S1.

Growth conditions and genomic DNA preparation

Necropsobacter massiliensis strain FF6^T (= CSUR P3511 = DSM 27814) was grown aerobically on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C. Bacteria grown on four petri dishes were resuspended in 5 × 100 µL of TE buffer; 150 µL of this suspension was diluted in 350 µL TE buffer

FIG. 3. Transmission electron microscopy of *Necropsobacter massiliensis strain* FF6^T strain. Cells are observed on Tecnai G20 transmission electron microscope operated at 200 keV. Scale bar = 500 µm.

10 × , 25 µL proteinase K and 50 µL sodium dodecyl sulfate for lysis treatment. This preparation was incubated overnight at 56°C. Extracted DNA was purified using three successive phenol–chloroform extractions and ethanol precipitation at −20°C of minimum 2 hours each. After centrifugation, the DNA was suspended in 65 µL EB buffer. The genomic DNA concentration was measured at 30.06 ng/µL using the Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA).

Genome sequencing and assembly

Genomic DNA of *Necropsobacter massiliensis* FF6^T was sequenced on the MiSeq sequencer (Illumina, San Diego, CA, USA) with the Paired-end and Mate-pair strategies. The paired-end and the mate-pair strategies were barcoded in order to be mixed respectively with ten other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina).

Genomic DNA was diluted to 1 ng/µL to prepare the paired-end library. The "tagmentation" step fragmented and tagged the DNA with an optimal size distribution at 1.5 kb. Then limited-cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter, Fullerton, CA, USA), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single-strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in single 39-hour runs at 2 × 250 bp.

Total information of 3.89 GB was obtained from a 416 k/mm^2 cluster density with a cluster passing quality control filters of 95.4% (7 899 000 clusters). Within this run, the index representation for *Necropsobacter massiliensis* was determined to

TABLE 2. **Differential characteristics of *Necropsobacter massiliensis* strain FF6^T with *Necropsobacter rosorum* [1], *Actinobacillus actinomycetemcomitans* [34,35], *Haemophilus influenzae* [34–36] and *Pasteurella multocida* [34–37]**

Character	*Necropsobacter massiliensis*	*Necropsobacter rosorum*	*Actinobacillus actinomycetemcomitans*	*Haemophilus influenzae*	*Pasteurella multocida*
Cell diameter (µm)	0.4	NA	0.5		
Gram stain	−	−	−	−	−
Motility	−	−	−	−	−
Endospore formation	−	NA	−	NA	NA
Production of					
Alkaline phosphatase	+	NA	+	+	Variable
Acid phosphatase	+	NA	+	NA	Variable
Catalase	−	+	+	+	+
Oxidase	+	+	−	+	+
β-Haemolysis	−	−	−	−	−
Urease	−	−	−	+	−
Indole	−	−	NA	+	+
Nitrate reductase	+	+	+	+	+
α-Galactosidase	−	+	NA	−	NA
β-Galactosidase					
α-glucosidase (PNPG)	+	+	NA	−	NA
β-Glucosidase	−	−	NA	−	NA
Esterase	+	NA	Variable	−	Variable
Esterase lipase	−	NA	Variable	NA	Variable
N-acetyl-β-glucosaminidase	−	NA	NA	NA	NA
Utilization of					
D-Fructose	+	−	+	NA	+
D-Mannose	+	+	+	NA	+
D-Xylose	+	+	Variable	+	+
D-Glucose	+	+	+	+	NA
Habitat	Human	Guinea pig	Human	Human	Human and animal

NA, data not available.

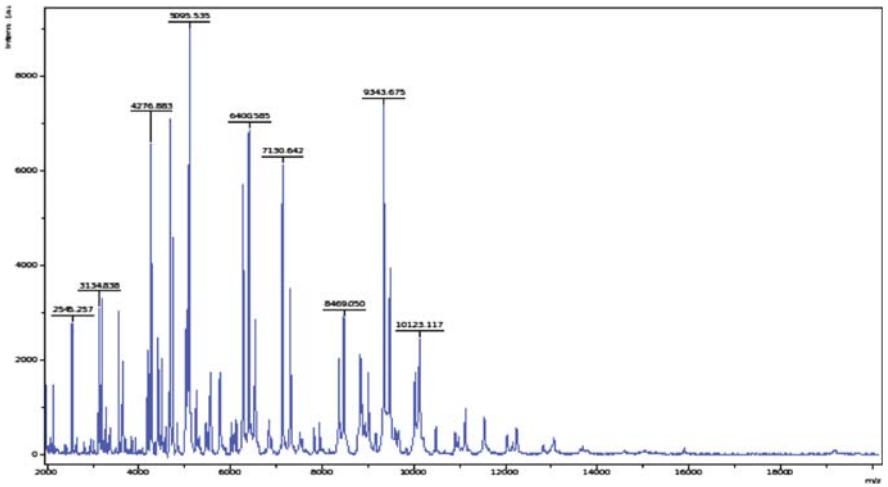


FIG. 4. Reference mass spectrum from *Necropsobacter massiliensis* strain FF6^T. Spectra from 12 individual colonies were compared and reference spectrum was generated.

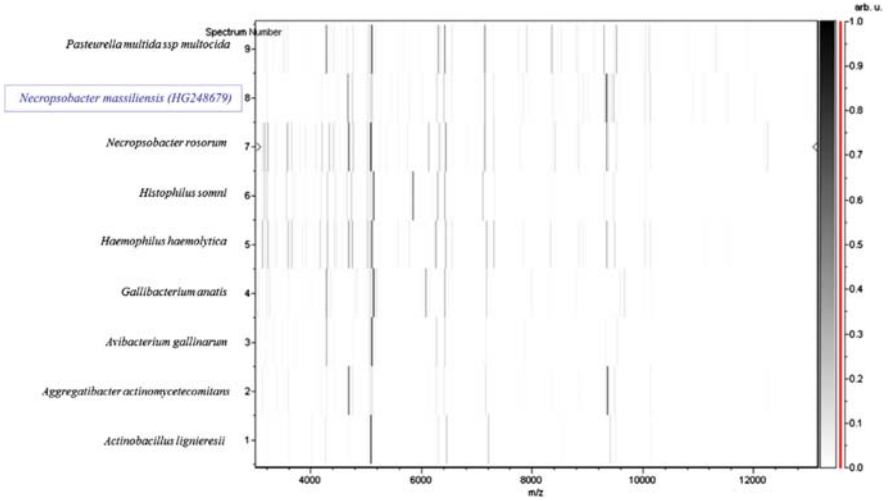


FIG. 5. Gel view comparing *Necropsobacter massiliensis* strain FF6^T to members of family *Pasteurellaceae*. Gel view displays raw spectra of all loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value; left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Colour bar and right y-axis indicating relation between colour peak is displayed; peak intensity indicated arbitrary units. Displayed species are indicated at left.

TABLE 3. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Paired end and mate pair 9 kb library
MIGS-29	Sequencing platforms	MiSeq
MIGS-31.2	Fold coverage	141×
MIGS-30	Assemblers	CLC 7
MIGS-32	Gene calling method	Prodigal
	Locus Tag	Not indicated
	GenBank ID	CDON00000000
	GenBank date of release	26 March 2015
	GOLD ID	Gp0102103
	BIOPROJECT	PRJEB4626
	Source material identifier	DSM 27814
	Project relevance	MALDI-TOF implementation in Dakar

MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

be 7.02% and to present 529 002 reads filtered according to the read qualities.

The mate-pair library was prepared with 1 µg of genomic DNA using the Nextera mate-pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1 kb up to 6.6 kb, with an optimal size at 4.5 kb. No size selection was performed, and 368 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments on the Covaris S2 device in microtubes (Covaris, Woburn, MA, USA).The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent) with an optimal peak at 672 bp. The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 10 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-hour run at 2 × 250 bp. *Necropsobacter massiliensis* strain FF6^T was determined to be 6.86%. The 639 775 reads were filtered according to the read qualities.

Genome annotation

Open reading frame (ORF) prediction was carried out using Prodigal [15] with default parameters. We removed the predicted ORFs if they spanned a sequencing gap region. Functional assessment of protein sequences was performed by comparing them with sequences in the GenBank [16] and the Clusters of Orthologous Groups (COGs) databases using BLASTP. tRNAs, rRNAs, signal peptides and transmembrane helices were identified using tRNAscan-SE 1.21 [17], RNAmmer [18], SignalP [19] and TMHMM [20], respectively. Artemis [21] was used for data management, whereas DNA Plotter [22] was used for visualization of genomic features. In-house perl and bash scripts were used to automate these routine tasks. ORFans were sequences which have no homology in a given database, *i.e.* in nonredundant (nr) or identified if their BLASTP E value was lower than 1e-03 for alignment lengths greater than 80 amino acids. PHAST was used to identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids [23].

To estimate the nucleotide sequence similarity at the genome level between *Necropsobacter massiliensis* and another ten members of the *Pasteurellaceae* family, we determined the average genomic identity of orthologous gene sequences (AGIOS) parameter as follows: orthologous proteins were detected using the Proteinortho software [24] (with the following parameters: E-value 1e-5, 30% percentage identity, 50% coverage and 50% algebraic connectivity) and genomes compared two by two. After fetching the corresponding nucleotide sequences of orthologous proteins for each pair of genomes, we determined the mean percentage of nucleotide sequence identity using the Needleman-Wunsch global alignment algorithm. Genomes from the genus *Necropsobacter* and closely related genera were used for the calculation of AGIOS values. The script created to calculate AGIOS values was named MAGi (Marseille Average genomic identity) and is written in perl and bioperl modules. GGDC analysis was also performed using the GGDC Web server (http://ggdc.dsmz.de) as previously reported [25,26].

Here, we compared the genome sequences of *Necropsobacter massiliensis* strain FF6^T (GenBank accession number CDON00000000) with those of *N. rosorum* strain P709T (CCMQ00000000), *Pasteurella multocida* subsp. *multocida* strain Pm70 (AE004439), *Haemophilus influenzae* strain Rd KW20 (L42023), *Haemophilus ducreyi* strain 35000HP (AE017143), *Histophilus somnus* strain 129PT (CP000436), *Haemophilus parasuis* strain SH0165 (CP001321), *Haemophilus parainfluenzae* strain T3T1 (FQ312002) and *Aggregatibacter aphrophilus* strain NJ8700 (CP001607).

Genome properties

The genome of *Necropsobacter massiliensis* strain FF6^T is 2 493 927 bp long with a 46.2% G+C content (Fig. 6). Of the 2363 predicted genes, 2309 were protein coding genes and 54 were RNA genes including 1 complete rRNA operon. A total of 1838 genes (77.7%) were assigned a putative function. A total of 210 were identified as ORFans (9.09%). The remaining genes were annotated as hypothetical proteins. The properties and the statistics of the genome are presented in Table 4. The distribution of genes into COGs functional categories is summarized in Table 5.

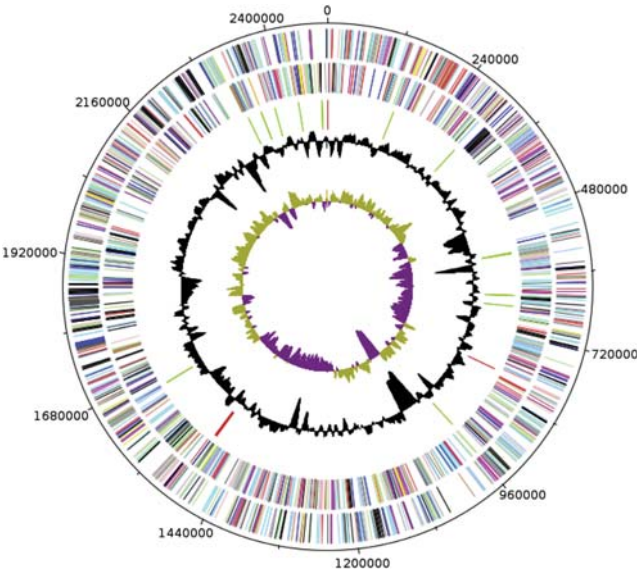


FIG. 6. Graphical circular map of *Necropsobacter massiliensis* strain FF6^T chromosome. From outside in, outer two circles shows open reading frames oriented in forward (coloured by COGs categories) and reverse (coloured by COGs categories) direction, respectively. Third circle marks indicate rRNA gene operon (green) and tRNA genes (red). Fourth circle shows G+C% content plot. Innermost circle shows GC skew, with purple indicating negative values and olive positive values.

Insights From Genome Sequence

Extended insights

The draft genome of *Necropsobacter massiliensis* (2.49 Mb) has a lower size than that of *N. rosorum* (2.52 Mb) but a larger size than those of *P. multocida* (2.25 Mb), *H. influenzae* (1.83 Mb), *H. ducreyi* (1.69 Mb), *H. somnus* (2.00 Mb), *H. parasuis* (2.26 Mb), *H. aphrophilus* (2.31 Mb) and *H. parainfluenzae* (2.08 Mb). The G+C content of *Necropsobacter massiliensis* (46.2%) was lower than that of *N. rosorum* (48.9%) but higher than those of *P. multocida* (40.40%), *H. influenzae* (38.15%), *H. ducreyi* (38.22%), *H. somnus* (37.20%), *H. parasuis* (39.99%), *H. aphrophilus* (42.23%) and *H. parainfluenzae* (39.57%). Because it has been suggested in the literature that the G+C content deviation is at most 1% within species, these data are an additional argument for the creation of a new taxon [27].

TABLE 4. **Genome information**

Attribute	Value	% of total
Genome size (bp)	2 493 927	
DNA coding (bp)	2 230 337	89.4
DNA G+C (bp)	1 151 339	46.2
DNA scaffolds	43	
Total genes	2363	100
Protein coding genes	2309	97.7
RNA genes	54	
Pseudo genes	Not indicated	
Genes in internal clusters	130	5.63
Genes with function prediction	1838	77.7
Genes assigned to COGs	2035	88.1
Genes with Pfam domains	75	3.24
Genes with signal peptides	210	9.09
Genes with transmembrane helices	561	24.3
CRISPR repeats	3	

COGs, Clusters of Orthologous Groups.

The protein-coding genes of *Necropsobacter massiliensis* is larger than those of *P. multocida, H. influenzae, H. ducreyi, H. somnus, H. parasuis, H. aphrophilus* and *H. parainfluenzae* (2012, 1603, 1717, 1791, 2021, 2218 and 1975, respectively) but smaller than that of *N. rosorum* (2311). However, the distribution of genes into categories was similar in all compared genomes. In addition, *Necropsobacter massiliensis* shared 2012, 1603, 1717, 1791, 2021, 1975, 2301 and 2218 orthologous genes with *P. multocida, H. influenzae, H. ducreyi, H. somnus, H. parasuis, H. aphrophilus* and *H. parainfluenzae,* respectively. Among species with standing in nomenclature, AGIOS values ranged from 66.32 between *N. rosorum* and *H. ducreyi* to 98.71% between *P. multocida* and *H. parainfluenzae* (Table 6). When

TABLE 5. Number of genes associated with general COGs functional categories[a]

Code	Value	Percentage	Description
J	152	6.58	Translation, ribosomal structure and biogenesis
A	1	0.04	RNA processing and modification
K	100	4.33	Transcription
L	127	5.50	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	26	1.13	Cell cycle control, cell division, chromosome partitioning
V	21	0.91	Defense mechanisms
T	29	1.26	Signal transduction mechanisms
M	117	5.07	Cell wall/membrane biogenesis
N	0	0.00	Cell motility
U	37	1.60	Intracellular trafficking and secretion
O	86	3.72	Posttranslational modification, protein turnover, chaperones
C	113	4.89	Energy production and conversion
G	182	7.88	Carbohydrate transport and metabolism
E	152	6.58	Amino acid transport and metabolism
F	55	2.38	Nucleotide transport and metabolism
H	86	3.72	Coenzyme transport and metabolism
I	44	1.91	Lipid transport and metabolism
P	112	4.85	Inorganic ion transport and metabolism
Q	8	0.35	Secondary metabolites biosynthesis, transport and catabolism
R	198	8.58	General function prediction only
S	172	7.45	Function unknown
—	197	8.53	Not in COGs

Total is based on total number of protein coding genes in annotated genome.
COGs, Clusters of Orthologous Groups.

compared to other species, *Necropsobacter massiliensis* exhibited AGIOS values ranging from 67.15 with *H. ducreyi* to 84.44 with *N. rosorum*. We obtained similar results using the GGDC software, as dDDH values ranged from 0.201 to 0.281 between studied species and were 0.275 between *N. rosorum*. These values confirm the status of *Necropsobacter massiliensis* as a new species.

Conclusions

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Necropsobacter massiliensis* sp. nov. that contains strain FF6^T. The strain was isolated from a cervical abscess of a 4-year-old Senegalese boy.

Taxonomic and nomenclatural proposals: description of *Necropsobacter massiliensis* strain FF6^T sp. nov.

Necropsobacter massiliensis (mas•il•ien'sis, L. gen., fem. n. *massiliensis*, of Massilia, the Latin name of Marseille, where this strain was characterized). On 5% sheep's blood–enriched Columbia agar (BioMérieux), colonies were 1 mm in diameter and grey. Cells are Gram negative and not motile, with a mean diameter of 0.4 µm (range, 0.2–0.6 µm) and a mean length of 1.5 µm (range, 0.9–2.1 µm). Catalase test was negative and oxidase test was positive. Positive reactions were observed for glycerol, ribose, D-xylose, D-mannose, D-glucose, inositol, N-acetyl glucosamine, D-fructose, D-maltose D-melibiose, D-trehalose, D-saccharose, D-raffinose, starch, potassium 5-ketogluconate, alkaline phosphatase, esterase, leucine arylamidase, phosphatase acid, α-glucosidase, naphthol-AS-BI-phosphohydrolase and α-glucosidase. Negative reactions were observed for D-mannitol, D-sorbitol, L-xylose, D-adonitol, methyl β-D-xylopyranoside, D-melezitose, inulin, α-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, N-acetyl-β-glucosaminidase, lipase, α-chrymotrypsin, cystine arylamidase, α-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase, N-acetyl-β-glucosaminidase, lipase, α-chrymotrypsin and cystine arylamidase.

Necropsobacter massiliensis strain FF6^T is susceptible to amoxicillin, amoxicillin/clavulanic acid, ceftriaxone, gentamicin, nitrofurantoin, rifampicin, trimethoprim/sulfamethoxazole and ciprofloxacin but resistant to erythromycin, doxycycline and vancomycin. The G+C content of the genome is 46.2%. The 16S rRNA and genome sequences of *N. massiliensis* strain FF6^T (= CSUR P3511 = DSM 27814) are deposited in GenBank under accession numbers HG428679 and CDON00000000, respectively. The type strain, FF6^T, was isolated from a cervical abscess of a 4-year-old Senegalese boy hospitalized in Hôpital Principal in Dakar, Senegal.

TABLE 6. Number of orthologous proteins shared between genomes (upper right) and AGIOS values obtained (lower left)

	N. massiliensis	*P. multocida.* subsp. *multocida*	*H. influenzae*	*H. ducreyi*	*H. somnus*	*H. parasuis*	*H. parainfluenzae*	*N. rosorum*	*A. aphrophilus*
N. massiliensis	2311[a]	71.37	70.61	67.15	71.73	68.01	71.29	84.44	72.12
P. multocida. subsp. *multocida*	1508	2012[a]	72.86	69.26	73.13	69.78	98.71	70.73	72.45
H. influenzae	1305	1271	1603[a]	69.89	72.96	70.15	72.81	69.71	73.69
H. ducreyi	1137	1107	1013	1717[a]	69.74	73.04	69.26	66.32	68.61
H. somnus	1306	1261	1124	1010	1791[a]	69.88	73.12	71.00	71.88
H. parasuis	1335	1290	1165	1068	1140	2021[a]	69.63	67.23	69.35
H. parainfluenzae	1523	1869	1275	1119	1270	1307	1975[a]	70.68	72.40
N. rosorum	1730	1514	1323	1111	1236	1335	1528	2301[a]	71.90
A. aphrophilus	1514	1407	1239	1065	1183	1204	1412	1463	2218[a]

Necropsobacter massiliensis FF6^T; *Necropsobacter rosorum*; *Pasteurella multocida* subsp. *multocida* Pm70; *Haemophilus influenzae* Rd KW20; *Haemophilus ducreyi* 35000HP; *Haemophilus somnus* 129PT; *Haemophilus parasuis* SH0165; HPI^g, *Haemophilus parainfluenzae* T3T1; *Aggregatibacter aphrophilus* NJ8700.
AGIOS, average genomic identity of orthologous gene sequences.
[a]Numbers of proteins per genome.

Acknowledgements

The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process. This study was funded by the Mediterranée-Infection Foundation. We thank J. Rathored for taking part in this work.

Conflict of Interest

None declared.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.nmni.2015.09.007.

References

[1] Christensen H, Korczak BM, Bojesen AM, Kuhnert P, Frederiksen W, Bisgaard M. Classification of organisms previously reported as the SP and Stewart-Letscher groups, with descriptions of *Necropsobacter* gen. nov. and of *Necropsobacter rosorum* sp. nov. for organisms of the SP group. Int J Syst Evol Microbiol 2011;61:1829–36.
[2] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 2014;42:D613–6.
[3] Tan TY, Ng LS, Eng LC. Five clinical cases of *Necropsobacter rosorum* bacteremia. J Clin Microbiol 2013;51:1337–40.
[4] Padmanabhan R, Robert C, Fenollar F, Raoult D, Fournier PE. Draft genome sequence of *Necropsobacter rosorum* strain P709T. Genome Announc 2014;2.
[5] Fall B, Lo CI, Samb-Ba B, et al. The ongoing revolution of MALDI-TOF mass spectrometry for microbiology reaches tropical Africa. Am J Trop Med Hyg 2015;92:641–7.
[6] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[7] Ramasamy D, Mishra AK, Lagier JC, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[8] Sentausa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect 2013;19:790–5.
[9] Reddy TBK, Thomas A, Stamatis D, et al. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification. Nucleic Acids Res 2014;43:1099–106.
[10] Lo CI, Mishra AK, Padhmanabhan R, et al. Non-contiguous finished genome sequence and description of *Clostridium dakarense* sp. nov. Stand Genomic Sci 2013;30:14–27.
[11] Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 2013;195:413–8.
[12] Seng P, Drancourt M, Gouriet F, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.
[13] Seng P, Rolain JM, Fournier PE, La Scola B, Drancourt M, Raoult D. MALDI-TOF–mass spectrometry applications in clinical microbiology. Future Microbiol 2010;5:1733–54.
[14] Field D, Garrity G, Gray T, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26:541–7.
[15] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.
[16] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:48–53.
[17] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[18] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucl Acids Res 2007;35:3100–8.
[19] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[20] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.
[21] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[22] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[23] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.
[24] Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.
[25] Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci 2010;2:142–8.
[26] Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence–based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013;14:60.
[27] Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 2014;64:352–6.
[28] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eukarya. Proc Natl Acad Sci U S A 1990;87:4576–9.
[29] Garrity GM, Bell JA, Lilburn T, Phylum XIV. *Proteobacteria* phyl. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. Bergeys's manual of systematic bacteriology (2nd ed., vol. 2, pt. B). New York: Springer; 2005. p. 1.
[30] Williams KP, Kelly DP. Proposal for a new class within the phylum *Proteobacteria, Acidithiobacillia* classis nov., with the type order *Acidithiobacillales*, and emended description of the class *Gammaproteobacteria*. Int J Syst Evol Microbiol 2013;63:2901–6.
[31] Christensen H, Kuhnert P, Busse HJ, Frederiksen WC, Bisgaard M. Proposed minimal standards for the description of genera, species and subspecies of the *Pasteurellaceae*. Int J Syst Evol Microbiol 2007;57:166–78.
[32] Pohl S. Family *Pasteurellaceae* fam. nov. in validation of the publication of new names and new combinations previously effectively published outside the IJSB, list no. 7. Int J Syst Bacteriol 1981;31:382–3.
[33] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
[34] Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol 1980;30:225–420.

[35] Nørskov-Lauritsen N, Kilian M. Reclassification of *Actinobacillus actinomycetemcomitans, Haemophilus aphrophilus, Haemophilus paraphrophilus and Haemophilus segnis* as *Aggregatibacter actinomycetemcomitans* gen. nov., comb. nov., *Aggregatibacter aphrophilus* comb. nov. and *Aggregatibacter segnis* comb. nov., and emended description of *Aggregatibacter aphrophilus* to include V factor–dependent and V factor–independent isolates. Int J Syst Evol Microbiol 2006;56:2135–46.

[36] Winslow CEA, Broadhurst J, Buchanan RE, Krumwiede C, Rogers LA, Smith GH. The families and genera of the bacteria preliminary report of the Committee of the Society of American Bacteriologists on characterization and classification of bacterial types. J Bacteriol 1917;2: 505–66.

[37] Mutters R, Ihm P, Pohl S, Frederiksen W, Mannheim W. Reclassification of the genus *Pasteurella* Trevisan 1887 on the basis of deoxyribonucleic acid homology, with proposals for the new species *Pasteurella dagmatis, Pasteurella canis, Pasteurella stomatis, Pasteurella anatis*, and *Pasteurella langaa*. Int J Syst Bacteriol 1985;35: 309–22.

Article 8:

« Non-contiguous finished genome sequence and description of Diaminobutyricimonas massiliensis strain FF2^T sp. nov. »

Cheikh Ibrahima Lo[1], Roshan Padhmanabhan[1], Oleg Madianikov[1], Alpha Kabinet Keita[1], Caroline Michelle[1], Jérôme Terras[1], Jaishriram Rathored[1], Ngor Faye[2], Didier Raoult[1, 3], Pierre-Edouard Fournier[1] and Florence Fenollar[1*]

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France Tel: +33 491 385 517
[2]Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann (Senegal)
[3]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Publié dans:

New Microbes and New Infections

November 2015, Volume 8, Pages 31-40

http://dx.doi.org/10.1016/j.nmni.2015.09.009

Résumé de l'article 8 : «Non-contiguous finished genome sequence and description of *Diaminobutyricimonas massiliensis* strain FF2^T sp. nov.»

La souche FF2^T a été isolée du sang d'un patient Sénégalais de 35 ans présentant de la fièvre inexpliquée dans le village de Dielmo au Sénégal. L'analyse par spectrométrie de masse de type MALDI-TOF de la souche FF2^T a généré des scores de 1.3 et 1.4 et n'a donc pas permis d'identification. L'analyse de la séquence de l'ARN 16S ribosomique de la souche FF2^T a montré une similarité de 97,47% avec la séquence de l'ARN 16S ribosomique de *Diaminobutyricimonas aerilata,* l'espèce officielle la plus proche phylogénétiquement.

Nous avons donc réalisé une étude polyphasique reposant sur des analyses phénotypiques et génomiques de la souche FF2^T. Celle-ci est une bactérie aérobie à Gram-négatif en forme de bâtonnet avec une longueur et un diamètre moyens de 1,3 et 0,5 µm, respectivement. La bactérie est aussi mobile. Les colonies ont un diamètre de 0,8 mm et une légère coloration jaune sur gélose Columbia enrichi avec 5% de sang de mouton. La croissance optimale est observée sur gélose après 48 heures d'incubation à 37°C en atmosphère aérobie supplémentée avec 5% de CO_2. La souche FF2^T se distingue des autres membres de la famille des *Microbacteriaceae* par la positivité de l'acide phosphatase et la négativité de la β-galactosidase et du mannose. La souche FF2^T présente des réactions positives pour la catalase, l'oxydase, la phosphatase alcaline, la N-acetyl-β-glucosaminidase, l'estérase, l'estérase-lipase, la leucine arylamidase, la phosphatase acide, l'α-glucosidase et l'esculine hydrolase. La souche FF2^T est *in vitro* sensible à la pénicilline, la

ceftriaxone, l'imipénème, la ciprofloxacine, la gentamicine, la rifampicine, la vancomycine et la doxycycline, mais résistante à l'érythromycine, la nitrofurantoïne, le métronidazole et le triméthoprime-sulfaméthoxazole.

Le génome de la souche $FF2^T$ compte 3 227 513 paires de base (1 chromosome pas de plasmide). Son taux de G+C est de 70,13%. Il contient 3 091 gènes codant pour des protéines et 56 gènes ARN dont 2 opérons ARN.

Sur la base de ces résultats, nous avons formellement proposé la création de *Diaminobutyricimonas massiliensis* sp. nov. dont la souche $FF2^T$ est la souche type.

Description de *Diaminobutyricimonas massiliensis* sp. nov.

Diaminobutyricimonas massiliensis (mas.si.li.e'n.sis. L. masc. n. *massiliensis*, de *Massilia*, le nom latin de Marseille, France, où la souche $FF2^T$ a été caractérisée). La souche type $FF2^T$ (= CSUR P3023 = DSM 27836) a été isolée du sang d'un patient Sénégalais de 35 ans avec une fièvre inexpliquée. Les séquences de l'ARN 16S ribosomique et du génome ont été déposées dans GenBank sous les numéros d'accession respectifs : HG315674 et CCSB00000000.

Noncontiguous finished genome sequence and description of *Diaminobutyricimonas massiliensis* strain FF2^T sp. nov.

C. I. Lo[1,2], R. Padhmanabhan[1], O. Mediannikov[1,2], A. K. Keita[1,2], C. Michelle[1], J. Terras[1], N. Faye[3], D. Raoult[1,2,4], P.-E. Fournier[1] and F. Fenollar[1,2]

1) *Institut hospitalo-universitaire Méditerranée-infection, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, Aix-Marseille Université, Marseille, France,* 2) *Campus International UCAD-IRD,* 3) *Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Dakar, Senegal and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Strain FF2^T was isolated from the blood sample of a 35 year-old febrile Senegalese male, in Dielmo, Senegal. This strain exhibited a 97.47% 16S rRNA sequence identity with *Diaminobutyricimonas aerilata*. The score from MALDI-TOF-MS does not allow any identification. Using a polyphasic study made of phenotypic and genomic analyses, strain FF2^T was Gram-negative, aerobic, motile, rod-shaped, and exhibited a genome of 3,227,513 bp (1 chromosome but no plasmid) with a G+C content of 70.13% that coded 3,091 protein-coding and 56 RNA genes. On the basis of these data, we propose the creation of *Diaminobutyricimonas massiliensis* sp. nov.

Keywords: Culturomics, *Diaminobutyricimonas massiliensis*, genome, taxono-genomics
Original Submission: 16 July 2015; **Revised Submission:** 7 September 2015; **Accepted:** 8 September 2015
Article published online: 16 September 2015

Corresponding author: F. Fenollar, Institut hospitalo-universitaire Méditerranée-infection, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, Aix-Marseille Université, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
E-mail: florence.fenollar@univ-amu.fr

Introduction

The *Microbacteriaceae* family is constituted of a large group of rod-shaped or, rarely, coccoid or mycelium-forming *Actinobacteria* [1,2]. In this group, there are 42 genera with validly published names [3]. Members of the family were found in diverse environments, including seawater, seaweed, soil, butter and cow feces but also in human specimens such as urine and human blood as well as in wounds [4–6]. *Diaminobutyricimonas aerilata* is the only species which has been officially validly published within the genus *Diaminobutyricimonas;* it was isolated from an air sample in Korea [7]. The genetic parameters used in the delineation of bacterial species include 16S rRNA sequence identity and phylogeny [8,9], genomic G+C content diversity and DNA-DNA hybridization [10]. These reference tools exhibit limitations, notably because their cutoff values vary across species or genera [11]. Thanks to the introduction of high-throughput sequencing techniques, more than 45 000 bacterial genomes have been fully sequenced and are currently available [12]. We recently proposed to incorporate genomic data into a polyphasic approach to describe new bacterial taxa (taxonogenomics). This strategy combines phenotypic characteristics, notably the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) spectrum, and genomic analysis and comparison [13].

Diaminobutyricimonas massiliensis strain FF2^T (= Collection de souches de l'Unité des Rickettsies (CSUR) P3023 = Deutsche Sammlung von Mikroorganismen (DSM) 27836) is designated as the type strain of *Diaminobutyricimonas massiliensis* sp. nov. This bacterium was isolated from a blood specimen from a 35-year-old febrile Senegalese patient as part of a study aiming at detecting bacterial pathogens associated with fever in malaria-negative patients [14–16].

Diaminobutyricimonas massiliensis is a Gram-positive, obligate aerobe, and motile rod-shaped bacterium.

Here we present a summary classification and a set of features of *Diaminobutyricimonas massiliensis* sp. nov. strain FF2^T (= CSUR P3023 = DSM 27836), together with the description of the complete genome sequencing and annotation. These characteristics support the circumscription of the species *Diaminobutyricimonas massiliensis.*

Organism Information

A blood sample was collected from a 35-year-old febrile Senegalese patient living in Dielmo, Senegal (Table 1). *Diaminobutyricimonas massiliensis* strain FF2^T (Table 1) was isolated

http://dx.doi.org/10.1016/j.nmni.2015.09.009

TABLE 1. Classification and general features of *Diaminobutyricimonas massiliensis* strain FF2^T [17]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain: *Bacteria*	TAS [31]
		Phylum: *Actinobacteria*	TAS [2]
		Class: *Actinobacteria*	TAS [32,33]
		Order: *Actinomycetales*	TAS [33,34]
		Family: *Microbacteriaceae*	TAS [35]
		Genus: *Diaminobutyricimonas*	TAS [7]
		Species: *Diaminobutyricimonas massiliensis*	IDA
		(Type) strain: FF2^T	IDA
	Gram stain	Positive	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Nonsporulating	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37°C	IDA
	pH range; optimum	7.2–6.9; 7.05	
	Carbon source	Unknown	NAS
MIGS-6	Habitat	Human blood	IDA
MIGS-6.3	Salinity	Unknown	
MIGS-22	Oxygen requirement	Aerobic	IDA
MIGS-15	Biotic relationship	Free-living	IDA
MIGS-14	Pathogenicity	Unknown	NAS
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection time	October 2012	IDA
MIGS-4.1	Latitude	13.716667	IDA
MIGS-4.1	Longitude	−16.416667	IDA
MIGS-4.4	Altitude	21 m above sea level	IDA

[a]Evidence codes: IDA, inferred from direct assay; TAS, traceable author statement (i.e. direct report exists in the literature); NAS, nontraceable author statement (i.e. not directly observed for the living, isolated sample but based on a generally accepted property for the species or anecdotal evidence). These evidence codes are from the Gene Ontology project (http://www.geneontology.org/GO.evidence.shtml) [36]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or by an expert mentioned in the acknowledgements.

in October 2012 in culture on agar enriched with 5% sheep's blood (bioMérieux, Marcy l'Etoile, France) in aerobic conditions after 48 hours' incubation at 37°C. Strain FF2^T exhibited a 97.47% 16S rRNA sequence identity with *Diaminobutyricimonas aerilata* (GenBank accession number JQ639052), the phylogenetically closest bacterial species with a validly published name (Fig. 1). The value was lower than the 98.8% 16S rRNA gene sequence threshold recommended by Meier-Kolthoff *et al.* [17] to delineate a new species within phylum *Actinobacteria* without carrying out DNA-DNA hybridization. Different growth temperatures (25°C, 30°C, 37°C, 45°C and 56°C) were tested. Growth occurred between 37°C and 45°C, but optimal growth was observed at 37°C after 48 hours' incubation in aerobic conditions. The colonies were 0.8 mm in diameter and exhibited a light yellow colour on agar enriched with 5% sheep's blood (bioMérieux). Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (bioMérieux) and under aerobic conditions, with or without 5% CO_2. Optimal growth was only obtained under aerobic conditions in the presence of 5% CO_2. No growth was obtained under anaerobic and microaerophilic conditions. The Gram staining was positive in fresh culture (less than 72 hours), but in old culture the bacteria may appear as a mixture of Gram-positive and Gram-negative bacteria rods (Gram variable) or as Gram-negative cells (Fig. 2). The motility test was positive. Cells grown on agar exhibited a mean diameter of 0.5 µm (range, 0.4–0.6 µm) and a mean length of 1.3 µm (range, 0.7–2 µm) under electron microscopy (Fig. 3).

Strain FF2^T was catalase and oxidase positive. Using an API 50CH strip (bioMérieux), a positive reaction was observed only for esculin ferric citrate. Negative reactions were observed for D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, D-trehalose, D-turanose, D-fucose, starch, glycogen, D-maltose, amygdalin, N-acetyl-glucosamine, methyl-αD-glucopyranoside and D-cellobiose. Using an API ZYM strip (bioMérieux), positive reactions were obtained for alkaline phosphatase, esterase, esterase-lipase, leucine arylamidase, acid phosphatase, N-acetyl-β-glucosaminidase and α-glucosidase. Negative reactions were observed for α-chymotrypsin, valine arylamidase, cystine arylamidase, trypsin, naphthol-AS-BI-phosphohydrolase, β-glucuronidase, β-glucosidase, α-galactosidase, β-galactosidase, α-mannosidase and α-fucosidase.

Diaminobutyricimonas massiliensis is susceptible to penicillin, ceftriaxone, imipenem, ciprofloxacin, gentamicin, rifampicin, vancomycin and doxycycline, but resistant to erythromycin, nitrofurantoin, metronidazole and trimethoprim/sulfamethoxazole. The phenotypic characteristics of *Diaminobutyricimonas massiliensis* strain FF2^T were compared to those of representative species from the family *Microbacteriaceae,* as summarized in Table 2.

MALDI-TOF protein analysis was carried out as previously described [18,19] using a Microflex LT (Brüker Daltonics,

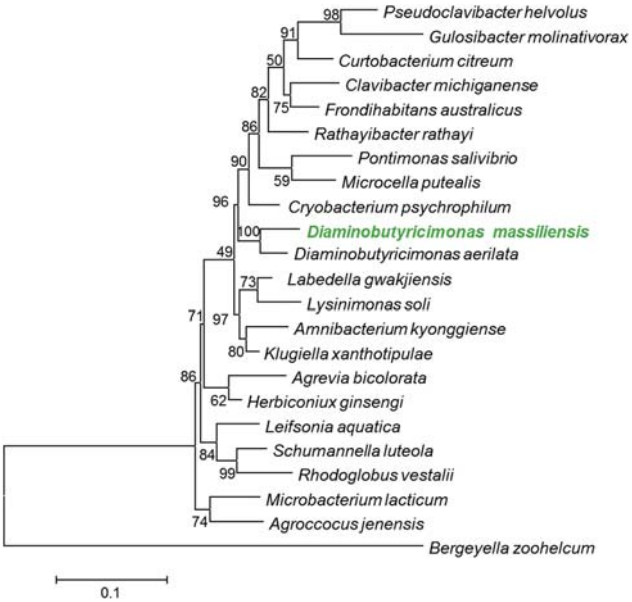


FIG. 1. Phylogenetic tree highlighting position of *Diaminobutyricimonas massiliensis* sp. nov., strain FF2^T, relative to other type strains within family *Microbacteriaceae*. Strains and their corresponding GenBank accession numbers for 16S rRNA are: *Pseudoclavibacter helvolus* strain DSM 20419, X77440; *Gulosibacter molinativorax* strain ON4, AJ306835; *Curtobacterium citreum* strain DSM 20528, X77436; *Clavibacter michiganensis* strain DSM 7483, X77434; *Frondihabitans australicus* strain E1HC-02, DQ525859; *Rathayibacter rathayi* strain DSM 7485, X77439; *Pontimonas salivibrio* strain CL-TW6, JQ639087; *Microcella putealis* strain CV2T, AJ717388; *Cryobacterium psychrophilum* strain DSM 4854, AJ544063; *Diaminobutyricimonas massiliensis* strain FF2^T, HG315674; *Diaminobutyricimonas aerilata* strain 6408J-67, JQ639052; *Labedella gwakjiensis* strain KSW2–17, DQ533552; *Lysinimonas soli* strain SGM3–12, JN378395; *Amnibacterium kyonggiense* strain KSL51201–037, FJ527819; *Klugiella xanthotipulae* strain 44C3, AY372075; *Agrevia bicolorata* strain VKM Ac-1804, AF159363; *Herbiconiux ginsengi* strain wged11, DQ473536; *Leifsonia aquatica* strain DSM 20146, D45057; *Schumannella luteola* strain KHIA, AB362159; *Rhodoglobus vestalii* strain LV3, AJ459101 *Microbacterium lacticum* strain DSM20427, X77441; *Agroccocus jenensis* strain DSM 9580, X92492; *Bergeyella zoohelcum* strain D658, M93153. Sequences were aligned with CLUSTALW, and phylogenetic inferences obtained using maximum-likelihood method within MEGA software. Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. *Bergeyella zoohelcum* strain D658 was used as outgroup. Scale bar = 10% nucleotide sequence divergence.

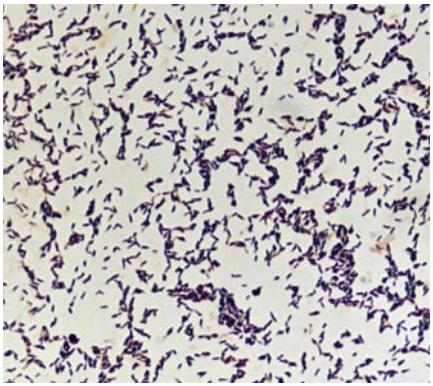


FIG. 2. Gram staining of *Diaminobutyricimonas massiliensis* strain FF2^T.

Leipzig, Germany). Twelve individual colonies were deposited on an MTP 384 MALDI-TOF target plate (Brüker). A total of 2 µL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 2.5% trifluoroacetic acid was distributed on each smear and air dried for 5 minutes at room temperature. The 12 individual spectra from strain FF2^T were imported into the MALDI BioTyper software (version 2.0; Brüker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6252 bacterial spectra. The scores previously established by Brüker Daltonics to validate whether the species could be compared to the database of the instrument were applied. Briefly, a score $\geq$2.000 with a species with a validly published name provided allows identification at the species level; a score of $\geq$1.700 and <2.000 allows identification at the genus level; and a score of <1.700 does not allow any identification. A score ranging between 1.3 and 1.4 was obtained for strain FF2^T and did not allow any identification. We added the spectrum from strain FF2^T to our database (Fig. 4). Finally, the gel view showed spectral differences with other members of the family *Microbacteriaceae* (Fig. 5).

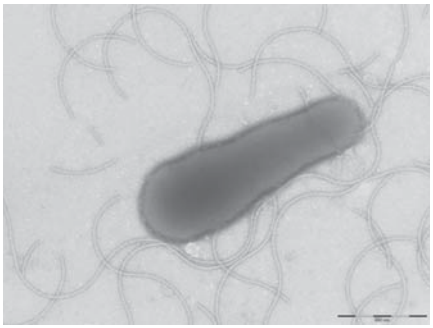


FIG. 3. Transmission electron microscopy of *Diaminobutyricimonas massiliensis* strain FF2^T. Cells are observed on Tecnai G20 operated at 200 keV. Scale bar = 500 nm.

Genome Sequencing Information

Genome project history

Strain FF2^T was selected for sequencing on the basis of its phylogenetic position and 16S rRNA similarity to other members of the family *Microbacteriaceae*. It was the first sequenced genome of a *Diaminobutyricimonas* species and the first genome of *Diaminobutyricimonas massiliensis* sp. nov. The GenBank accession number is CCSB00000000 and consists of 39 large contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [20].

Growth conditions and DNA isolation

Diaminobutyricimonas massiliensis sp. nov., strain FF2^T (= CSUR P3023 = DSM 27836), was grown on Columbia agar enriched with 5% sheep's blood (bioMérieux) at 37°C in an aerobic atmosphere. Bacteria grown on five plates were collected and resuspended in 500 µL of Tris-EDTA buffer 10×. A total of 100 µL of this solution was then completed with 400 µL TE buffer 10×, 25 µL proteinase K, and 50 µL sodium dodecyl sulfate, then incubated overnight at 56°C for complete cells lysis. The next day, DNA was extracted using a phenol-chloroform protocol. The final DNA extract was resuspended in 65 µL EB buffer. The genomic DNA concentration was measured at 35.8 ng/µL using the Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA).

Genome sequencing and assembly

Genomic DNA of *Diaminobutyricimonas massiliensis* was sequenced on the MiSeq sequencer (Illumina, San Diego, CA, USA) using two sequencing strategies: paired end and mate pair. The paired-end and mate-pair strategies were barcoded in order to be mixed, respectively, with ten other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina).

The genomic DNA was diluted to 1 ng/µL to prepare the paired-end library. The tagmentation step fragmented and tagged the DNA with an optimal size distribution at 4.5 kb. Then

TABLE 2. Differential phenotypic characteristics of *Diaminobutyricimonas massiliensis* strain FF2^T (data from this study), *Labedella gwakjiensis* strain KSW2–17 [37], *Diaminobutyricimonas aerilata* strain 6408J-67 [7], *Lysinimonas soli* strain SGM3–12 [38] and *Cryobacterium psychrophilum* strain DSM 4854 [39].

Property	*D. massilensis*	*D. aerilata*	*L. gwakjiensis*	*L. soli*	*C. psychrophilum*
Cell diameter (µm)	0.4–0.6	0.5–0.6	0.3–0.4	0.4–0.5	0.5–0.7
Gram stain	+	+	+	+	+
Motility	+	+	–	–	–
Endospore formation	–	–	–	–	–
Oxygen requirement	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic
Production of:					
Alkaline phosphatase	+	+	–	–	NA
Acid phosphatase	+	–	–	–	NA
Catalase	+	–	+	–	+
Oxidase	+	–	–	+	NA
Nitrate reductase	–	–	–	–	–
Urease	–	–	–	–	NA
α-Galactosidase	–	–	+	+	NA
β-Galactosidase	–	+	+	+	NA
β-Glucuronidase	–	–	–	–	NA
α-Glucosidase	+	+	+	+	NA
β-Glucosidase	–	+	+	–	NA
Esterase	+	–	–	+	NA
Esterase lipase	+	+	+	+	NA
Acid from:					
Ribose	–	–	–	–	–
Mannose	–	+	+	+	+
D-Fructose	–	+	+	NA	+
D-Glucose	–	+	+	+	+
Habitat	Human blood	Air sample	Dried seaweed	Soil	Soil

NA, data not available.

limited-cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter, Fullerton, CA, USA), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single-strand library was loaded onto the reagent cartridge and then onto the instrument, along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in a single 39-hour run at 2 × 250 bp.

Total information of 3.89 Gb was obtained from a 654 K/mm^2 cluster density with a cluster passing quality control filters of 93.7% (12 204 000 clusters). Within this run, the index

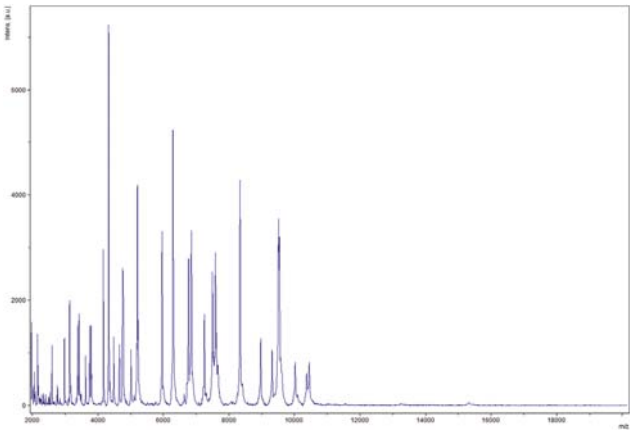


FIG. 4. Reference mass spectrum from *Diaminobutyricimonas massiliensis* strain FF2^T. This reference spectrum was generated by comparison of 12 individual colonies.

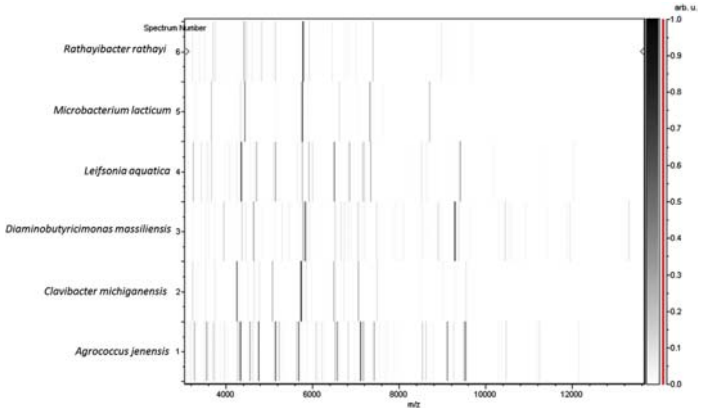


FIG. 5. Gel view comparing *Diaminobutyricimonas massiliensis* strain FF2^T spectrum to other members of family *Microbacteriaceae*. Gel view displays raw spectra of all loaded spectrum files arranged in pseudo-gel-like look. X-axis records m/z value. Left y-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Colour bar and right y-axis indicate relation between colour peak is displayed; peak intensity in arbitrary units. Displayed species are indicated at left.

representation for *Diaminobutyricimonas massiliensis* was determined to 4.72%. The 539 968 reads were filtered according to the read qualities.

The mate-pair library was prepared with 1 µg of genomic DNA using the Nextera mate-pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1 kb up to 6.6 kb, with an optimal size at 4.5 kb. No size selection was performed, and 368 ng of tagmented fragments was circularized. The circularized DNA was mechanically sheared to small fragments on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies) with an optimal peak at 504 bp. The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 10 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-hour run at 2 × 250 bp. *Diaminobutyricimonas massiliensis* was determined to 9.49%. The 1 884 171 reads were filtered according to the read qualities.

TABLE 3. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Paired end and mate pair 9 kb library
MIGS-29	Sequencing platform	Illumina MiSeq
MIGS-31.2	Fold coverage	95×
MIGS-30	Assemblers	Newbler version 2.5.3
MIGS-32	Gene calling method	Prodigal
	Locus tag	Not indicated
	GenBank ID	CCSB00000000
	GenBank date of release	8 September 2013
	GOLD ID	Gp0102104
	BioProject ID	PRJEB4274
MIGS-13	Source material identifier	DSM 27836
	Project relevance	Study of causes of nonmalarial fever

Genome annotation

Open reading frame (ORF) prediction was performed using Prodigal [21] with default parameters. We removed the predicted ORFs if they spanned a sequencing gap region. Functional assessment of protein sequences was performed by comparing them with sequences in the GenBank [22], and Clusters of Orthologous Groups (COGs) databases using BLASTP. tRNAs, rRNAs, signal peptides and transmembrane helices were identified using tRNAscan-SE 1.21 [23], RNAmmer [24], SignalP [25] and TMHMM [26], respectively. Artemis [27] was used for data management, while DNA Plotter [28] was used for visualization of genomic features. In-house perl and bash scripts were used to automate these routine tasks. ORFans were sequences which have no homology in a given database, i.e. in a nonredundant (nr) or identified if their BLASTP E value was lower than $1e^{-03}$ for alignment lengths greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E value of $1e^{-05}$.

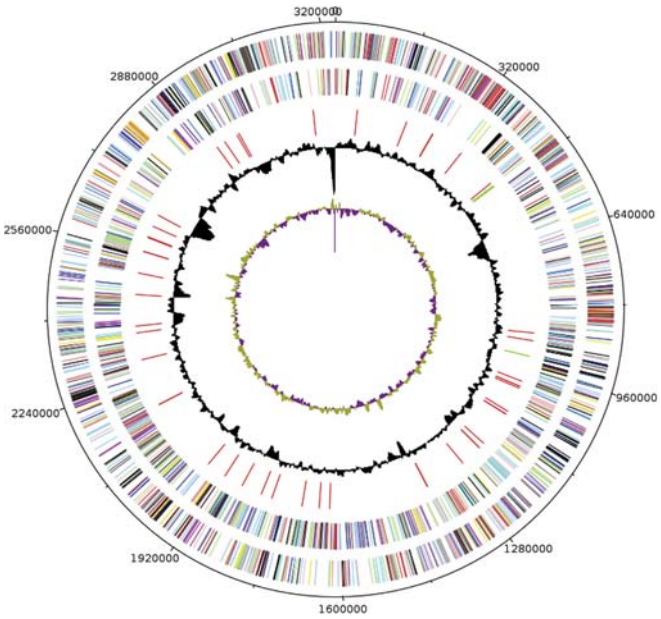
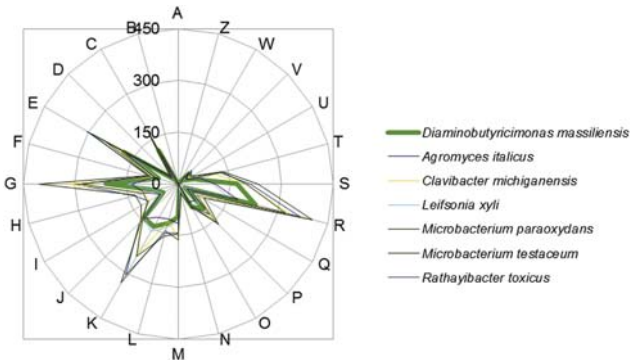


FIG. 6. Distribution of functional classes of predicted genes of *Diaminobutyricimonas massiliensis* strain FF2^T along with other genomes belonging to *Microbacteriaceae* family according to clusters of orthologous groups of proteins.

FIG. 7. Graphical circular map of *Diaminobutyricimonas massiliensis* strain FF2 ^{T} chromosome. From outside in, outer two circles show open reading frames oriented in forward (coloured by COGs categories) and reverse (coloured by COGs categories) directions, respectively. Third circle marks rRNA gene operon (green) and tRNA genes (red). Fourth circle shows G+C% content plot. Innermost circle displays GC skew, with purple and olive indicating negative and positive values, respectively.



PHAST was used to identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids [29].

To estimate the nucleotide sequence similarity at the genome level between *Diaminobutyricimonas massiliensis* and another six members of the family of *Microbacteriaceae*, we determined the average genomic identity of orthologous gene sequences (AGIOS) parameter as follows: orthologous proteins were detected using the Proteinortho software (with the following parameters: E value $1e^{-5}$, 30% identity, 50% coverage and algebraic connectivity of 50%) [30] and genomes compared two by two (Fig. 6). After fetching the corresponding nucleotide sequences of orthologous proteins for each pair of genomes, we determined the mean percentage of nucleotide sequence identity using the Needleman-Wunsch global alignment algorithm. The script created to calculate AGIOS values was named MAGi (Marseille Average genomic identity) and was written in perl and bioperl modules.

Genome properties

The genome of *Diaminobutyricimonas massiliensis* strain FF2^T is 3 227 513 bp long with a 70.13% G+C content (Fig. 7). Of the 3147 predicted genes, 3091 were protein-coding genes and 56 were RNAs. A total of 1937 genes (62.66%) were assigned a putative function. The properties and statistics of the genome are summarized in Table 4. The distribution of genes into COGs functional categories is presented in Table 5.

Insights From the Genome Sequence

Extended insights

We compared the genome from *Diaminobutyricimonas massiliensis* strain FF2^T to those of other members of the *Microbacteriaceae* family, including *Agromyces italicus* strain DSM 16388 (GenBank accession number ATXF00000000), *Clavibacter michiganensis* strain NCPPB 382 (AM711867), *Leifsonia xyli* strain CTCB07 (AE016822), *Microbacterium paraoxydans* strain 77MFTsu3.2 (AQYI00000000), *Microbacterium testaceum* strain StLB037 (AP012052) and *Rathayibacter toxicus* strain DSM 7488 (AUDF00000000). The draft genome of *Diaminobutyricimonas massiliensis* (3.22 Mb) is larger than that of *L. xyli* and *R. toxicus* (2.58 and 2.3 Mb, respectively) but is smaller than those of *C. michiganensis, M. paraoxydans* and *M. testaceum* (3.39, 3.47 and 3.98 Mb, respectively). The G+C content of *Diaminobutyricimonas massiliensis* (70.13%) is higher than those of *L. xyli, M. paraoxydans* and *R. toxicus* (67.70%, 69.50% and 61.50%, respectively) but lower than those of *A. italicus, M. testaceum* and *C. michiganensis* (70.2%, 70.30% and 72.53%, respectively). The gene content of *Diaminobutyricimonas massiliensis* (3147 genes) is larger than that of *L. xyli* and *R. toxicus* (2373 and 2188 genes, respectively) and smaller than those of *C. michiganensis, A. italicus, M. paraoxydans* and *M. testaceum* (3167, 3399, 3403 and 3727 genes, respectively). The AGIOS values between *D. massiliensis* and other members of the family *Microbacteriaceae* ranged from 68.85% with *R. toxicus* to 73.55% with *C. michiganensis* (Table 6).

TABLE 4. Nucleotide content and gene count levels of genome

Attribute	Value	% Of total[a]
Genome size (bp)	3 227 513	
DNA coding (bp)	2 925 192	90.63
DNA G+C (bp)	2 263 455	70.13
DNA scaffolds	ND	
Total genes	3147	100
Protein-coding genes	3091	98.22
RNA genes	56	
Pseudo genes	31	1.0
Gens in internal clusters	215	6.955
Genes with function prediction	1937	62.66
Genes assigned to COGs	2233	72.24
Genes with Pfam domains	172	5.56
Genes with peptide signals	186	6.017
Genes with transmembrane helices	846	27.36
ORFan genes	780	25.23
CRISPR repeats	4	

COGs, Clusters of Orthologous Groups; ND, not determined.
[a]Total is based on either genome size (in base pairs) or total number of protein coding genes in annotated genome.

TABLE 5. Number of genes associated with general COGs functional categories

Code	Value	Percentage[a]	Description
J	141	4.56	Translation, ribosomal structure and biogenesis
A	1	0.03	RNA processing and modification
K	142	4.59	Transcription
L	118	3.82	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	18	0.58	Cell cycle control, cell division, chromosome partitioning
V	42	1.36	Defense mechanisms
T	51	1.65	Signal transduction mechanisms
M	97	3.14	Cell wall/membrane biogenesis
N	13	0.42	Cell motility
U	21	0.68	Intracellular trafficking and secretion
O	78	2.52	Posttranslational modification, protein turnover, chaperones
C	112	3.62	Energy production and conversion
G	210	6.79	Carbohydrate transport and metabolism
E	183	5.92	Amino acid transport and metabolism
F	59	1.91	Nucleotide transport and metabolism
H	72	2.33	Coenzyme transport and metabolism
I	50	1.62	Lipid transport and metabolism
P	103	3.33	Inorganic ion transport and metabolism
Q	23	0.74	Secondary metabolites biosynthesis, transport and catabolism
R	233	7.54	General function prediction only
S	170	5.50	Function unknown
—	296	9.57	Not in COGs

COGs, Clusters of Orthologous Groups.
[a]Total is based on total number of protein coding genes in annotated genome.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Diaminobutyricimonas*

TABLE 6. Numbers of orthologous proteins shared between genomes (upper right) and AGIOS values (lower left)

	D. massiliensis	*A. italicus*	*C. michiganensis*	*L. xyli*	*M. paraoxydans*	*M. testaceum*	*R. toxicus*
D. massiliensis	3091[a]	72.99	73.55	72.96	70.61	70.83	68.85
A. italicus	1405	3376[a]	73.34	73.70	71.66	71.86	68.98
C. michiganensis	1347	1356	2983[a]	73.21	71.29	71.34	68.98
L. xyli	1045	1081	1072	2028[a]	70.61	70.97	69.34
M. paraoxydans	1351	1442	1397	1059	3514[a]	75.30	67.08
M. testaceum	1368	1515	1455	1074	1687	3676[a]	67.24
R. toxicus	1050	1068	1105	953	1072	1119	2188[a]

AGIOS, average genomic identity of orthologous gene sequences.
Diaminobutyricimonas massiliensis strain FF2^T; *Agromyces italicus* strain DSM 16388; *Clavibacter michiganensis* strain NCPPB; *Leifsonia xyli* strain CTCB07; *Microbacterium paraoxydans* strain 77MFTsu3.2; *Microbacterium testaceum* strain StLB037; *Rathayibacter toxicus* strain DSM 7488.
[a]Numbers of proteins per genome.

massiliensis sp. nov. that contains the strain FF2^T as type strain. This bacterial strain was isolated from a blood specimen from a 35-year-old Senegalese man with unexplained fever.

Description of *Diaminobutyricimonas massiliensis* sp. nov.

Diaminobutyricimonas massiliensis (mas•si•li•e'n.sis. L. masc. n. *massiliensis*, of Massilia, the Latin name for Marseille, France, where type strain FF2^T was characterized).

Colonies are 0.8 mm in diameter and light yellow on agar enriched with 5% sheep's blood. Cells are Gram positive, rod shaped, motile and aerobic, with a mean diameter and length of 0.5 and 1.3 µm, respectively. Optimal growth was observed after 48 hours on agar enriched with 5% sheep's blood at 37°C in an aerobic atmosphere supplemented with 5% CO_2. *Diaminobutyricimonas. massiliensis* is positive for catalase, oxidase, alkaline phosphatase, N-acetyl-β-glucosaminidase, esterase, esterase-lipase, leucine arylamidase, acid phosphatase, α-glucosidase and esculin ferric citrate. Negative reactions were observed for D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, D-trehalose, D-turanose, D-fucose, starch, glycogen, D-maltose, amygdalin, N-acetyl-glucosamine, methyl-αD-glucopyranoside, D-cellobiose, α-chymotrypsin, valine arylamidase, cystine arylamidase, trypsin, naphthol-AS-BI-phosphohydrolase, β-glucuronidase, β-glucosidase, α-galactosidase, β-galactosidase, α-mannosidase and α-fucosidase. *Diaminobutyricimonas massiliensis* is susceptible to penicillin, ceftriaxone, imipenem, ciprofloxacin, gentamicin, rifampicin, vancomycin and doxycycline but resistant to erythromycin, nitrofurantoin, metronidazole and trimethoprim/sulfamethoxazole. The G+C content of the genome is 70.13%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers HG315674 and CCSB00000000, respectively.

The type strain FF2^T (= CSUR P3023 = DSM 27836) was isolated from the blood of a Senegalese man with unexplained fever.

Conflict of Interest

None declared.

Acknowledgements

The authors thank the Xegen company (http://www.xegen.fr/) for automating the genomic annotation process. This study was funded by the Mediterranean Infection Foundation.

References

[1] Park YH, Suzuki KI, Yim DG, et al. Suprageneric classification of peptidoglycan group B actinomyces by nucleotide sequencing of 5S ribosomal RNA. Antonie Van Leeuwenhoek 1993;64:307–2313.
[2] Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. Int J Syst Bacteriol 1997;47:479–91.
[3] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 2014;42:D613–6.
[4] Aravena-Román M, Inglis TJ, Siering C, Schumann P, Yassin AF. *Canibacter oris* gen. nov., sp. nov., isolated from an infected human wound. Int J Syst Evol Microbiol 2014;64:1635–40.
[5] Fang H, Lv W, Huang Z, Liu SJ, Yang H. *Gryllotalpicola reticulitermitis* sp. nov. isolated from a termite gut. Int J Syst Evol Microbiol 2014;3.
[6] Lin YC, Uemori K, De Briel DA, Arunpairojana V, Yokota A. *Zimmermannella helvola* gen. nov., sp. nov., *Zimmermannella alba* sp. nov., *Zimmermannella bifida* sp. nov., *Zimmermannella faecalis* sp. nov. and *Leucobacter albus* sp. nov., novel members of the family *Microbacteriaceae*. Int J Syst Evol Microbiol 2004;54:1669–76.
[7] Jang YH, Kim SJ, Hamada M, et al. *Diaminobutyricimonas aerilata* gen. nov., sp. nov., a novel member of the family *Microbacteriaceae* isolated from an air sample in Korea. J Microbiol 2012;50:1047–52.
[8] Ramasamy D, Mishra AK, Lagier JC, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[9] Sentausa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect 2013;19:790–5.
[10] Rossello-Mora R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E,

editor. Molecular identification, systematics, and population structure of prokaryotes. Berlin: Springer; 2006. p. 23–50.
[11] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[12] Reddy TBK, Thomas A, Stamatis D, et al. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification. Nucleic Acids Res 2014;43:1099–106.
[13] Lo CI, Mishra AK, Padhmanabhan R, et al. Non contiguous finished genome sequence and description of *Clostridium dakarense* sp. nov. Stand Genomic Sci 2013;9:14–27.
[14] Sokhna C, Mediannikov O, Fenollar F, et al. Point-of-care laboratory of pathogen diagnosis in rural Senegal. PLoS Negl Trop Dis 2013;7: e1999.
[15] Mediannikov O, Socolovschi C, Edouard S, et al. Common epidemiology of *Rickettsia felis* infection and malaria, Africa. Emerging Infect Dis 2013;19:1775–83.
[16] Fenollar F, Mediannikov O, Socolovschi C, et al. *Tropheryma whipplei* bacteremia during fever in rural West Africa. Clin Infect Dis 2010;51: 515–21.
[17] Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 2013;195:413–8.
[18] Seng P, Drancourt M, Gouriet F, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.
[19] Fall B, Lo CI, Samb-Ba B, et al. The ongoing revolution of MALDI-TOF mass spectrometry for microbiology reaches tropical Africa. Am J Trop Med Hyg 2015;92:641–7.
[20] Field D, Garrity G, Gray T, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26: 541–7.
[21] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.
[22] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:48–53.
[23] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[24] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.
[25] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: signalp 3.0. J Mol Biol 2004;340:783–95.
[26] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.
[27] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[28] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[29] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39:347–52.
[30] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: de-tection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 2011;12:124.
[31] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. Proc Natl Acad Sci U S A 1990;87:4576–9.
[32] Euzéby JP, Tindall BJ. Nomenclatural type of orders: corrections necessary according to Rules 15 and 21a of the Bacteriological Code (1990 Revision) and designation of appropriate nomenclatural types of classes and subclasses. Request for an opinion. Int J Syst Evol Microbiol 2001;51:725–7.
[33] Zhi XY, Li WJ, Stackebrandt E. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *Actinobacteria*, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. Int J Syst Evol Microbiol 2009;59:589–608.
[34] Garrity GM, Holt JG. The road map to the manual. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's manual of systematic bacteriology (2nd ed., vol. 1). New York: Springer; 2001. p. 119–69.
[35] Validation of the publication of new names and new combinations previously effectively published outside the IJSB List No. 53. Int J Syst Bacteriol 1995;45:418–9.
[36] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
[37] Lee SD. *Labedella gwakjiensis* gen. nov., sp. nov., a novel actinomycete of the family *Microbacteriaceae*. Int J Syst Evol Microbiol 2007;57:2498–502.
[38] Jang YH, Kim SJ, Tamura T, et al. *Lysinimonas soli* gen. nov., sp. nov., isolated from soil, and reclassification of *Leifsonia kribbensis Dastager* et al 2009 as *Lysinimonas kribbensis* sp. nov., comb. nov. Int J Syst Evol Microbiol 2013;63:1403–10.
[39] Suzuki K, Sasaki J, Uramoto M, Nakase T, Komagata K. *Cryobacterium psychrophilum* gen. nov., sp. nov., nom. rev., comb. nov., an obligately psychrophilic actinomycete to accommodate "*Curtobacterium psychrophilum*" Inoue and Komagata 1976. Int J Syst Bacteriol 1997;47:474–8.

Article 9:

« High-quality genome sequence and description of Bacillus ndiopicus strain FF3T sp. nov.»

Cheikh Ibrahima Lo[1, 2], Roshan Padhmanabhan[1], Oleg Mediannikov[1,2], Aurelia Caputo[1], Caroline Michelle[1], Ngor Faye[3], Cheikh Sokhna[2], Didier Raoult[1, 2, 4], Pierre-Edouard Fournier[1] and Florence Fenollar[1, 2*]

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de Médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France Tel: +33 491 385 517,
[2]Campus international UCAD-IRD, Dakar, Senegal
[3]Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann, Senegal
[4]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Publié dans :

New Microbes and New Infections

November 2015Volume 8, Pages 154–163

http://dx.doi.org/10.1016/j.nmni.2015.10.009

Résumé de l'article 9: ***«High-quality genome sequence and description of Bacillus ndiopicus strain FF3T sp. nov.»***

La souche $FF3^T$ a été isolée de la flore cutanée d'un homme Sénégalais de 39 ans vivant dans le village de Ndiop au Sénégal. L'analyse par spectrométrie de masse MALDI-TOF n'a pas permis d'identification. Une analyse moléculaire a montré que cette souche était une nouvelle espèce du genre *Bacillus*, avec une similarité de séquence de l'ARN 16S ribosomique de 96,8% avec *Bacillus massiliensis.* La souche $FF3^T$ est Gram positive, anaérobie facultative et en forme de bâtonnet. Son génome fait 4 068 720 paires de bases avec 3 982 protéines-codantes, 67 gènes ARN et un contenu G+C de 37,03%.

La croissance optimale est observée à 37°C en atmosphère aérobie enrichi avec 5% de CO_2. La bactérie possède des flagelles donc mobile. Les colonies mesurent 1 mm de diamètre sur gélose Columbia enrichie avec 5% de sang de mouton. *Bacillus ndiopicus* est Gram positif, anaérobie facultative, catalase positive et oxydase négative.

B. ndiopicus est *in vitro* sensible à la pénicilline, la ceftriaxone, l'imipenème, la rifampicine, l'amoxicilline-acide clavulanique, la ciprofloxacine, la gentamicine et la doxycycline. Par contre cette bactérie est résistante au métronidazole et à la nitrofurantoïne. En outre, des réactions positives ont été observées avec le citrate, l'alcaline phosphatase, l'estérase, la lipase, et l'alpha-chymotrypsine.

Description de *Bacillus ndiopicus sp. nov.*

B. ndiopicus (n.dio.pi.cus. L. gen. masc. n. *ndiopicus*, de Ndiop, nom du village dans lequel le prélèvement cutané de la souche $FF3^T$ a été réalisé). La souche type $FF3^T$ (= CSUR P3025 = DSM 27837) a été isolée de la flore cutanée d'un homme Sénégalais de 39 ans en bonne santé. Les séquences de l'ARN 16S ribosomal et du génome ont été déposées dans GenBank avec les numéros d'accession respectifs: HG315675 et CCAP000000000.

High-quality genome sequence and description of *Bacillus ndiopicus* strain FF3^T sp. nov.

C. I. Lo[1,2], R. Padhmanabhan[1], O. Mediannikov[1,2], A. Caputo[1], C. Michelle[1], N. Faye[3], C. Sokhna[2], D. Raoult[1,2,4], P.-E. Fournier[1] and F. Fenollar[1,2]

1) *Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, Marseille, France,* 2) *Campus International UCAD-IRD, Dakar,* 3) *Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Fann, Senegal and* 4) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Strain FF3^T was isolated from the skin-flora of a 39-year-old healthy Senegalese man. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry did not allow any identification. This strain exhibited a 16S rRNA sequence similarity of 96.8% with *Bacillus massiliensis*, the phylogenetically closest species with standing nomenclature. Using a polyphasic study made of phenotypic and genomic analyses, strain FF3^T was Gram-positive, aeroanaerobic and rod shaped and exhibited a genome of 4 068 720 bp with a G+C content of 37.03% that coded 3982 protein-coding and 67 RNA genes (including four rRNA operons). On the basis of these data, we propose the creation of *Bacillus ndiopicus* sp. nov.

Keywords: *Bacillus ndiopicus*, genome, Senegal, skin, taxonogenomics
Original Submission: 11 August 2015; **Revised Submission:** 7 October 2015; **Accepted:** 9 October 2015
Article published online: 19 October 2015

Corresponding author: F. Fenollar, Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France
E-mail: florence.fenollar@univ-amu.fr

Introduction

Bacillus subtilis was the first type species described in the genus *Bacillus* (Cohn 1872) [1]. Currently there are 301 species and seven subspecies with validly published names [2]. Generally members of this genus are environmental bacteria present in soil, food, and fresh and sea water. In humans, some strains can be pathogenic, such as *Bacillus cereus* (associated mainly with food poisoning) and *Bacillus anthracis* (the causative agent of anthrax) [3–5]. Other strains are saprophytes [6]. Several *Bacillus* species are also isolated from different plants in which they are endophytes [7].

Recently high-throughput genome sequencing and mass spectrometry analyses of bacteria have given unprecedented access to an abundance of genetic and proteomic information [8–10]. Currently a polyphasic approach is performed to describe new bacterial taxa, including their genome sequence, matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) spectrum, and major phenotypic characteristics such as Gram staining, culture, metabolic characteristics, habitat and (if applicable) pathogenicity [9,10].

Bacillus ndiopicus strain FF3^T (= CSUR P3025 = DSM 27837) is designated as the type strain of *Bacillus ndiopicus*. This bacterium is a Gram-positive rod that is aeroanaerobic. This bacterium was isolated from the skin of a healthy Senegalese man as part of a culturomics [11] study aiming at cultivating bacterial species from skin flora.

Here we provide a summary classification and set of features for *B. ndiopicus* sp. nov. strain FF3^T, together with the description of the complete genomic sequencing and annotation. These characteristics support the circumscription of the species *B. ndiopicus*.

Organism information

Classification and features

In December 2012, a skin specimen was sampled with a swab from a healthy Senegalese volunteer living in Ndiop, a rural village in the Guinean–Sudanian area in Senegal (Table 1). This 39-year-old man was included in a research project approved by the National Ethic Committee for health research (CNERS) in Senegal and the ethics committee of the Institut Fédératif de Recherche IFR48, Faculty of Medicine, Marseille, France (agreements 09-022 and 11-017) [12].

Strain FF3^T (Table 1) was isolated by cultivation on 5% blood's sheep enriched Columbia agar (bioMérieux, Marcy l'Etoile, France), under aerobic conditions, in December 2012.

http://dx.doi.org/10.1016/j.nmni.2015.10.009

TABLE 1. Classification and general features of *Bacillus ndiopicus* strain FF3T [15]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain: *Bacteria*	TAS [27]
		Phylum: *Firmicutes*	TAS [28,29]
		Class: *Bacilli*	TAS [30,31]
		Order: *Bacillales*	TAS [32]
		Family: *Bacillaceae*	TAS [33]
		Genus: *Bacillus*	TAS [34,35]
		Species: *Bacillus ndiopicus*	IDA
		(Type) strain: FF3T	IDA
	Gram stain	Positive	IDA
	Cell shape	Rods	IDA
	Motility	Motile	IDA
	Sporulation	Sporulating	NAS
	Temperature range	Mesophile	IDA
	Optimum temperature	37°C	IDA
	pH range; optimum	5.6–8.4; 7.0	IDA
	Carbon source	Unknown	
MIGS-6	Habitat	Human skin	IDA
MIGS-6	Salinity	Unknown	
MIGS-22	Oxygen requirement	Aeroanaerobic	IDA
MIGS-15	Biotic relationship	Free-living	IDA
MIGS-14	Pathogenicity	Unknown	
MIGS-4	Geographic location	Ndiop, Senegal	TAS
MIGS-5	Sample collection	December 2012	TAS
MIGS-4.1	Latitude	14.5333	TAS
MIGS-4.1	Longitude	−16.2667	TAS
MIGS-4.4	Altitude	5 m above sea level	TAS

MIGS, minimum information about a genome sequence.
[a]Evidence codes are as follows: IDA, inferred from direct assay; TAS, traceable author statement (i.e., a direct report exists in the literature); NAS, nontraceable author statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species or anecdotal evidence). These evidence codes are from the Gene Ontology project (http://www.geneontology.org/GO.evidence.shtml) [36]. If the evidence code is IDA, then the property should have been directly observed, for the purpose of this specific publication, for a live isolate by one of the authors, or an expert or reputable institution mentioned in the acknowledgements.

B. ndiopicus strain FF3T exhibited a 96.8% nucleotide sequence similarity with *Bacillus massiliensis* (Glazunova *et al.*, 2006), the phylogenetically closest *Bacillus* species (Fig. 1). These values were lower than the 98.7% 16S rRNA gene sequence threshold recommended by Meier-Kolthoff *et al.* [13] to delineate a new species within the phylum *Firmicutes* without carrying out DNA-DNA hybridization. Different growth temperatures (25, 30, 37, 45 and 56°C) were tested. Optimal growth was observed at 37 and 45°C after 24 hours of incubation; weak growth was noticed at 30°C. Colonies were 1 mm in diameter and transparent on 5% blood-enriched Columbia agar. Growth of the strain was tested under anaerobic and microaerophilic conditions using the GENbag anaer and GENbag microaer systems, respectively (bioMérieux), and under aerobic conditions, with or without 5% CO_2. Optimal growth was obtained under aerobic condition with 5% CO_2 and under microaerophilic condition at 37 and 45°C.

Gram staining showed Gram-positive rods (Fig. 2). The motility test was positive by means of peritrichous flagella. Cells grown on agar have a mean diameter of 1.2 µm (ranging from 0.8 to 1.6 µm) and a mean length of 2.5 µm (ranging from 1.8 to 3.2 µm) (Fig. 3).

Strain FF3T exhibited catalase and oxidase activities. Using the API ZYM strip (bioMérieux), positive reactions were observed with alkaline phosphatase, esterase, α-chymotrypsin and lipase. Negative reactions were observed for leucine arylamidase, valine arylamidase, cystine arylamidase, phosphatase acid, trypsin, naphthol-AS-BI-phosphohydrolase, β-glucuronidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Using the API 20E strip (bioMérieux), only the citrate test was positive; all others tests were negative, including indole, β-galactosidase, urease, ornithine decarboxylase, mannitol, sorbitol and rhamnose fermentation. Using the API 50CH strip (bioMérieux), no positive reaction was observed, including for glycerol, D-arabinose, D-xylose, L-rhamnose, amygdalin, D-cellobiose, D-fucose, potassium 5-ketogluconate, L-arabitol, starch, D-maltose and D-mannose. *B. ndiopicus* was susceptible *in vitro* to penicillin, amoxicillin, amoxicillin−clavulanic acid, ceftriaxone, imipenem, gentamicin, ciprofloxacin, erythromycin, doxycycline, rifampicin and vancomycin, but resistant to nitrofurantoin and metronidazole. When compared with representative species from the genus *Bacillus*, *B. ndiopicus* strain FF3T exhibited several phenotypic differences, which are summarized in Table 2.

MALDI-TOF protein analysis was performed using a Microflex LT (Bruker Daltonics, Leipzig, Germany), as previously reported [14]. The scores previously established by Bruker allowing validating (or not) the identification of species compared to the database of the instrument were applied. Briefly, a score of $\geq$2.000 with a species with a validly published name provided allows the identification at the species level; a score of $\geq$1.700 and <2.000 allows the identification at the genus level; and a score of <1.700 does not allow any identification. We performed 12 distinct deposits from 12 isolated colonies of strain FF3T. Two microliters of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 2.5% trifluoroacetic acid were distributed on each smear and submitted at air drying for 5 minutes. Then the spectra from the 12 different colonies were imported into MALDI Biotyper 2.0 software (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6252 bacterial spectra including 199 spectra from 104 *Bacillus* species. Scores ranging from 1.2 to 1.4 were obtained for strain FF3T, suggesting that this isolate was not a member of any known species. The reference mass spectrum from strain FF3T was incremented in our database (Fig. 4). The gel view highlighted spectrum differences with other *Bacillus* species (Fig. 5).

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its 16S rRNA similarity, phylogenetic position and phenotypic

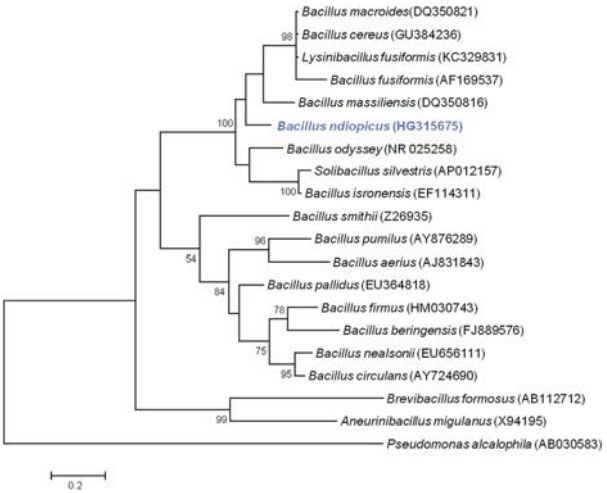


FIG. 1. Phylogenetic tree highlighting the position of *Bacillus ndiopicus* strain FF3^T relative to the most closely related type strains within the genus *Bacillus*. The strains and their corresponding GenBank accession numbers for 16S rRNA genes are provided (type = ^{T}), and in parentheses we indicate GA if the genome is available or GNA if the genome is not available at the National Center for Biotechnology Information website: *Bacillus macroides* strain LMG 18474 (GNA), *Bacillus cereus* strain ZQN6, *Lysinibacillus fusiformis* strain H1k (GA: AYMK00000000), *Bacillus massiliensis* strain 4400831 (GA: JPVQ00000000), *Bacillus ndiopicus* strain FF3^T (GA: CCAP000000000), *Bacillus odyssey* strain NBRC 100172 (GA: JPVP00000000), *Solibacillus silvestris* strain StLB046 (GA: AP012157), *Bacillus isronensis* (GA: AMCK00000000), *Bacillus smithii* strain 7_3_47FAA (GA: ACWF00000000), *Bacillus pumilus* strain BA06 (GA: AMDH00000000), *Bacillus aerius* strain 24K, *Bacillus pallidus* strain CW 7, *Bacillus firmus* strain DS1 (GA: APVL00000000), *Bacillus beringensis* strain BR035 (GNA), *Bacillus nealsonii* strain AAU1 (GA: ASRU00000000), *Bacillus circulans* NBRC 13626 (GNA), *Brevibacillus formosus* strain F12 (GNA), *Aneurinibacillus migulanus* (GA: GCA_000878905), and *Pseudomonas alcaliphila* strain JAB1 (GNA). Sequences were aligned using MUSCLE [40] and the phylogenetic tree inferred by the maximum likelihood method with Kimura two-parameter model from MEGA6 software [41]. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. *P. alcaliphila* was used as outgroup. Scale bar = rate of substitution per site of 0.2.

differences with other members of the genus *Bacillus*, which support that *Bacillus ndiopicus* strain FF3^T likely represents a new bacterial species. This strain is part of a study aiming to characterize the skin flora of healthy Senegalese people. Currently there are more of 270 sequenced genomes of *Bacillus* species [8]. Strain FF3^T is the first genome of *B. ndiopicus* sp. nov., and its GenBank accession number is CCAP000000000. The genome consists of 23 large contigs. Table 3 shows the project information and its association with minimum information about a genome sequence (MIGS)

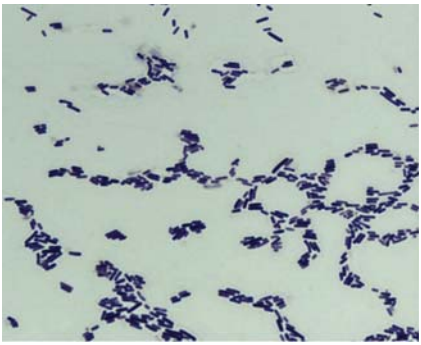


FIG. 2. Gram staining of *Bacillus ndiopicus* strain FF3^T.

2.0 compliance [15]; associated MIGS records are summarized.

Growth conditions and DNA isolation

Bacillus ndiopicus strain FF3^T (= CSUR P3025 = DSM 27837) was grown aerobically on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C. Then we suspended all bacterial colonies in 500 µL of Tris-EDTA (TE) buffer 10×. We remove 100 µL of this solution. This volume is completed by 400 µL TE buffer 10×, 25 µL proteinase K and 50 µL sodium dodecyl sulfate and then incubated overnight at 56°C for complete cells lysis. The next day this lysate is purified by washing with a phenol–chloroform solution three times. It is precipitated in absolute ethanol and incubated at −20°C for at least 2 hours. After a first centrifugation at 4°C for 30 minutes at 8000 rpm, the pellet is taken up in 70% ethanol kept at −20°C. A second centrifugation in the same conditions for 20 minutes is performed. After drying the tube in an oven at 37°C for 5 minutes, the DNA is taken up with 65 µL with buffer EB. The genomic DNA concentration was measured at 47.7 ng/µL by the Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA).

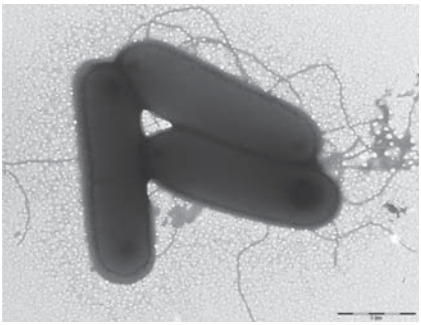


FIG. 3. Transmission electron microscopy of *Bacillus ndiopicus* strain FF3^T. Cells were observed on a Tecnai G20 device operated at 200 keV. Scale bar = 1 µm.

Genome sequencing and assembly

Genomic DNA of *Bacillus ndiopicus* was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) with two applications, paired end and mate pair. The paired-end and the mate-pair strategies were barcoded in order to be mixed with 11 other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina).

The genomic DNA was diluted to 1 ng/µL to prepare the paired-end library. The tagmentation step fragmented and tagged the DNA with an optimal size distribution at 0.95 kb. Then limited-cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter, Fullerton, CA, USA), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in a single 39-hour run in 2 × 250 bp.

Total information of 6.8 Gb was obtained from a 807K/mm^2 cluster density, with a cluster passing quality control filters of 90.88% (14 553 000 clusters). Within this run, the index representation for *Bacillus ndiopicus* was determined to 17.96% and present 2 375 297 reads filtered according to the read qualities.

The mate-pair library was prepared with 1 µg of genomic DNA using the Nextera mate-pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments were ranged in size from 1.5 to 13 kb, with an optimal size at 8 kb. No size selection was performed, and 600 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments on a Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent). The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 10 pM, the pool of libraries was loaded onto the reagent

TABLE 2. Differential characteristics of *Bacillus ndiopicus* strain FF3^T with *B. kribbensis* [37], *B. massiliensis* [38], *B. vireti* [39], *B. soli* [39]

Property	*B. ndiopicus*	*B. kribbensis*	*B. massiliensis*	*B. vireti*	*B. soli*
Cell diameter (μm)	0.8–1.6	1.4–2.0	0.3–0.5	0.6–0.9	0.6–1.2
Oxygen requirement	Aeroanaerobic	Aerobic	Aerobic	Facultative anaerobic	Facultative anaerobic
Gram stain	+	+	–	–	Variable
Motility	+	+	+	+	+
Endospore formation	+	+	+	+	+
Production of:					
Alkaline phosphatase	+	NA	NA	NA	NA
Acid phosphatase	–	NA	NA	NA	NA
Catalase	+	+	+	NA	NA
Oxidase	–	–	+	NA	NA
Nitrate reductase	–	–	–	+	+
Urease	–	NA	+	–	–
α-Galactosidase	–	NA	NA	NA	NA
β-Galactosidase	–	NA	NA	NA	NA
β-Glucuronidase	–	+	NA	NA	NA
α-Glucosidase	–	+	NA	NA	NA
β-Glucosidase	–	+	NA	NA	NA
Esterase	+	+	NA	NA	NA
Esterase lipase	+	+	NA	NA	NA
Naphthol-AS-BI-phosphohydrolase	–	+	NA	NA	NA
N-acetyl-β-glucosaminidase	–	NA	NA	+	+
Utilization of:					
5-Keto-gluconate	–	NA	–	–	–
D-Xylose	–	+	–	–	–
D-Fructose	–	+	–	+	+
D-Glucose	–	+	–	+	+
D-Mannose	–	–	–	+	+
Habitat	Human skin	Soil	Human CSF	Soil	Soil

+, positive result; –, negative result; CSF, cerebrospinal fluid; NA, data not available.

cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 42-hour run in 2 × 250 bp. *Bacillus ndiopicus* was determined to 8.09%. The 1 023 790 reads were filtered according to the read qualities. CLC Genomics Workbench 8.5.x was used for genome assembly.

Genome annotation

Open reading frames (ORFs) prediction was carried out using Prodigal [16] with default parameters. We removed the predicted ORFs if they spanned a sequencing gap region. Functional assessment of protein sequences was performed by comparing them with sequences in the GenBank [17] and Clusters of

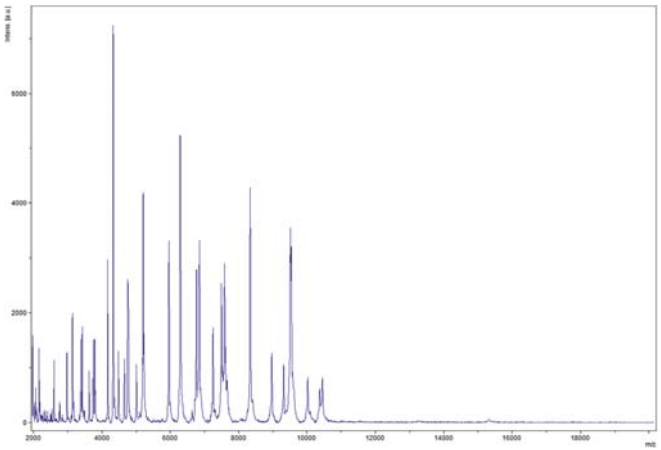


FIG. 4. Reference mass spectrum from *Bacillus ndiopicus* strain FF3^T. Spectra from 12 individual colonies were compared and reference spectrum generated.

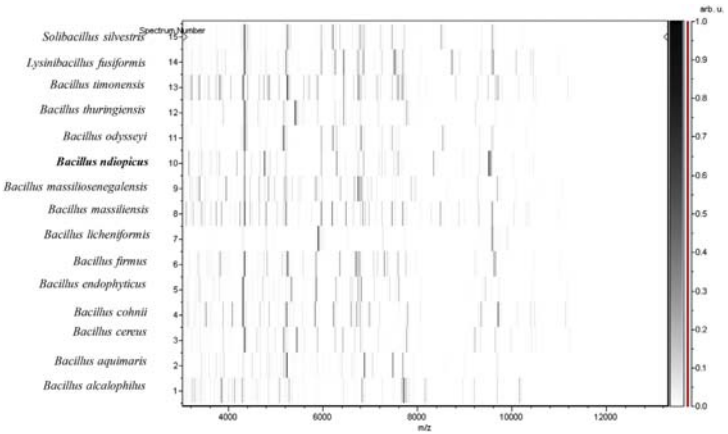


FIG. 5. Gel view comparing *Bacillus ndiopicus* strain FF3T spectrum to other members of family *Bacillaceae*. Gel view displays raw spectra of all loaded spectrum files arranged in pseudo-gel-like look. The *x*-axis records *m*/*z* value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Color bar and right *y*-axis indicating relation between color peak is displayed, with peak intensity in arbitrary units. Displayed species are indicated at left.

Orthologous Groups (COGs) databases using BLASTP. tRNAs, rRNAs, signal peptides and transmembrane helices were identified using tRNAscan-SE 1.21 [18], RNAmmer [19], SignalP [20] and TMHMM [21], respectively. Artemis [22] was used for data management, and DNA Plotter [23] was used for visualization of genomic features. In-house Perl and bash scripts were used to automate these routine tasks. ORFans were sequences which have no homology in a given database—that is, nonredundant (nr) or identified if their BLASTP *E* value was lower than 1e-03 for alignment lengths greater than 80 aa. PHAST was used to identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids [24].

TABLE 3. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Paired end and mate pair
MIGS-29	Sequencing platforms	MiSeq
MIGS-31.2	Fold coverage	52×
MIGS-30	Assemblers	CLC genomics workbench
MIGS-32	Gene calling method	Prodigal
	Locus tag	Not reported
	GenBank ID	CCAP000000000
	GenBank date of release	March 18, 2014
	GOLD ID	Gp0101144
	BIOPROJECT	PRJNA224116
MIGS-13	Source material identifier	DSM 27837
	Project relevance	Study of human skin flora

MIGS, minimum information about a genome sequence.

To estimate the nucleotide sequence similarity at the genome level between *B. ndiopicus* and other members of *Bacillaceae* family, orthologous proteins were detected by Proteinortho software [25] (with the following parameters: *E* value 1e-5, 30% percentage of identity, 50% coverage and algebraic connectivity of 50%) and genomes compared two by two. After fetching the corresponding nucleotide sequences of orthologous proteins for each pair of genomes, we determined the mean percentage of nucleotide sequence identity using the Needleman-Wunsch global alignment algorithm. The script created to calculate AGIOS (average genomic identity of orthologous gene sequences) values was named MAGi (Marseille Average genomic identity) and is written in Perl and Bioperl modules.

Genome properties

The genome of *B. ndiopicus* strain FF3T is 4 068 720 bp long (one chromosome, no plasmid) with a 37.03% G+C content (Fig. 6). Of note, we acknowledge the fact that because the genome of *Bacillus ndiopicus* is a draft sequence, its exact size might be slightly different from that of our sequence, but given the fold coverage (52×), we are confident that the missing fragments are probably small and do not significantly influence the genome size. Of the 3982 predicted genes, 3915 were protein-coding genes and 67 were RNAs. A total of 1697 genes (43.34%) were assigned a putative function. The properties of the genome are presented in Table 4. Using PHAST software,

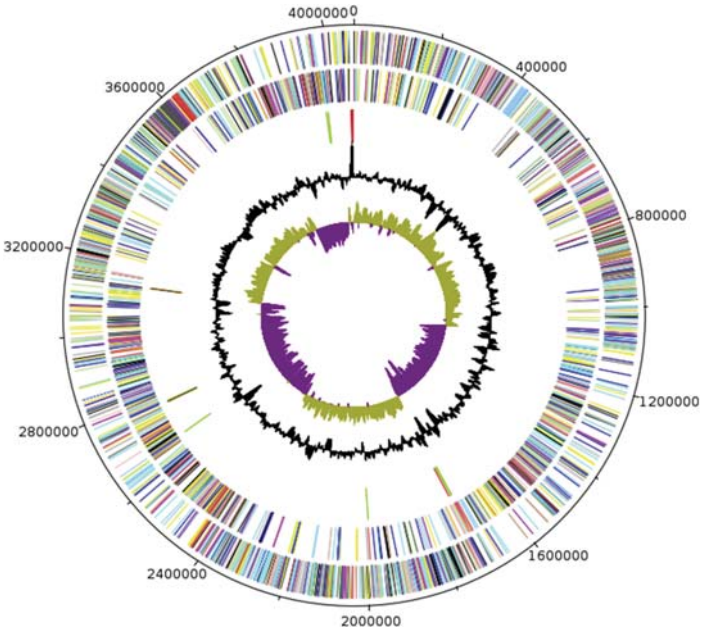


FIG. 6. Graphical circular map of *Bacillus ndiopicus* strain FF3^T chromosome. From outside in, outer two circles show ORFs oriented in forward (colored by COGs categories) and reverse (colored by COGs categories) directions, respectively. Third circle marks rRNA gene operon (red) and tRNA genes (green). Fourth circle shows G+C% content plot. Innermost circle shows GC skew; purple and olive indicate negative and positive values, respectively.

three prophage regions were identified, including one complete and two incomplete prophages (Table 5). A total of 167 were identified as ORFans (42.65%). The distribution of genes into COGs functional categories is presented in Table 6.

Genomic comparative

Today there are more than 277 sequenced genomes of *Bacillus* species (finished and draft) available in Genomes Online Database [3]. Here we compared *B. ndiopicus* genome sequence against other members of genus *Bacillus*, including *Bacillus coagulans* strain 2-6, *B. coagulans* strain 36D1, *Lysinibacillus sphaericus* strain C3-41, *Bacillus bataviensis* stain LMG 21833, and *Bacillus isronensis* strain B3W22. Table 7 shows a comparison of genome size, G+C% content, and number of proteins for each genome selected for taxonogenomic study. Indeed, *Bacillus ndiopicus* has a genome size of 4.06 Mb higher than those of *B. coagulans* 2–6 (3.07 Mb), *B. coagulans* 36D1 (3.55 Mb) and *B. isronensis* B3W22 (4.02 Mb) but lower than those of *B. bataviensis* LMG 21833 (5.37 Mb) and *Lysinibacillus sphaericus* C3-41 (4.82 Mb).

Bacillus ndiopicus strain FF3^T has a G+C content (37.03%) lower than those of all the compared species such as *B. coagulans* strain 2-6 (47.3%), *B. coagulans* strain 36D1 (46.5%), *B. bataviensis* strain LMG 21833 (39.6%), *B. isronensis* strain

TABLE 4. Genome information

Attribute	Value	% of total[a]
Genome size (bp)	4 068 720	
DNA coding (bp)	3 460 992	85.0
DNA G+C (bp)	1 506 586	37.03
DNA scaffolds	8	
Total genes	3982	100
Protein coding genes	3915	98.31
RNA genes	67	
Pseudo genes	51	1.18
Genes in internal clusters	208	4.82
Genes with function prediction	1697	43.34
Genes assigned to COGs	1892	48.32
Genes with Pfam domains	3235	75.45
Genes with peptide signals	60	1.53
Genes with transmembrane helices	530	13.5
CRISPR	4	

COGs, Clusters of Orthologous Groups database; CRISPR, clustered regularly interspaced short palindromic repeat.
[a]Total is based on total number of protein-coding genes in annotated genome.

TABLE 5. Identified prophage regions of *Bacillus ndiopicus*[a]

Region	Region length (kb)	Completeness	No. of coding sequence	Region position	Phage	GC%
1	15.6	Incomplete	16	269 940–285 579	PHAGE_Geobac_virus_E2_NC_009552	36.36
2	62.1	Complete	82	1 127 027–1 189 204	PHAGE_Thermu_OH2_NC_021784	37.40
3	18.7	Incomplete	25	1 843 157–1 861 873	PHAGE_Clostr_phiC2_NC_009231	36.67

[a]Region indicates number assigned to region; region length, length of sequence of that region (in bp); completeness, prediction of whether region contains a complete or incomplete prophage; region position, start and end positions of region on bacterial chromosome; phage, phage with highest number of proteins most similar to those in region; and GC%, percentage of GC nucleotides of region.

B3W22 (38.8%) and *L. sphaericus* strain C3-41 (37.1%). As it has been suggested in the literature that the G+C content deviation is at most 1% within species, these data are an additional argument for the creation of a new taxon [26].

The number of orthologous genes shared between *B. ndiopicus* and other *Bacillus* species as well as the average percentage nucleotide identity calculated using the MAGi method is tabulated in Table 8. On the basis of the analysis of MAGi, the AGIOS ranged from 61.79 to 95.94% among the studied members. The range of AGIOS calculated using MAGi varies from 61.79 to 70.95% between *B. ndiopicus* and other compared *Bacillus* species. Antibiotic resistance genes were detected within the genome using the ARDB website (Table 9).

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses (taxonogenomics), we formally propose the creation of *Bacillus ndiopicus* sp. nov. that contains strain FF3^T as the type strain.

TABLE 6. Number of genes associated with general COGs functional categories

Code	Value	%	Description
J	166	4.24	Translation, ribosome structure and biogenesis
A	0	0.00	RNA processing and modification
K	231	5.90	Transcription
L	127	3.24	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	33	0.84	Cell cycle control, cell division, chromosome partitioning
V	76	1.94	Defense mechanisms
T	126	3.21	Signal transduction mechanisms
M	112	2.86	Cell wall/membrane biogenesis
N	23	0.58	Cell motility
U	21	0.53	Intracellular trafficking and secretion
O	65	1.66	Posttranslational modification, protein turnover, chaperones
C	105	2.68	Energy production and conversion
G	98	2.50	Carbohydrate transport and metabolism
E	231	5.90	Amino acid transport and metabolism
F	75	1.91	Nucleotide transport and metabolism
H	89	2.27	Coenzyme transport and metabolism
I	70	1.78	Lipid transport and metabolism
P	155	3.95	Inorganic ion transport and metabolism
Q	24	0.61	Secondary metabolites biosynthesis, transport and metabolism
R	348	8.88	General function prediction only
S	303	7.73	Function unknown
—	195	4.98	Not in COGs

COGs, Clusters of Orthologous Groups database.

The strain was isolated from the skin of a 39-year-old healthy Senegalese man living in Ndiop, Senegal.

Description of *Bacillus ndiopicus* strain FF3^T sp. nov.

B. ndiopicus (n.dio.pi.cus. L. gen. masc. n. *ndiopicus*, of Ndiop, the name of the Senegalese village where the man from whom strain FF3^T was cultivated lives).

Cells stain Gram positive, are rod shaped and endospore forming, motile and have a mean diameter of 1.2 µm and a mean length of 2.5 µm. Peritrichous flagellae were observed. Colonies are 1 mm in diameter and transparent on 5% sheep's blood. Optimal growth is achieved at 37°C in an aerobic atmosphere supplemented with 5% CO_2. Catalase and oxidase activities are positive. Positive reactions were obtained with citrate, alkaline phosphatase, esterase, lipase and α-chymotrypsin. Negative reactions were observed for leucine arylamidase, valine arylamidase, cystine arylamidase, phosphatase acid, trypsin, naphthol-AS-BI-phosphohydrolase, β-glucuronidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. *B. ndiopicus* is susceptible *in vitro* to penicillin, amoxicillin, amoxicillin–clavulanic acid, ceftriaxone, imipenem, gentamicin, ciprofloxacin, erythromycin, doxycycline, rifampicin and vancomycin, but resistant to nitrofurantoin and metronidazole.

TABLE 7. Genome comparison of *Bacillus ndiopicus* strain FF3^T with other *Bacillus* species

No.	Organism	Accession	Size (Mb)	No. of proteins	GC %
1	*Bacillus coagulans* 2-6	NC_015634	3.07	2971	47.3
2	*Bacillus coagulans* 36D1	NC_016023	3.55	3289	46.5
3	*Lysinibacillus sphaericus* C3-41	CP000817	4,82	4584	37.1
4	*Bacillus bataviensis* LMG 21833	NZ_AJLS00000000	5.37	5207	39.6
5	*Bacillus isronensis* B3W22	NZ_AMCK01000000	4.02	3883	38.8
6	*Bacillus ndiopicus* strain FF3^T	CCAP000000000	4.06	3915	37.03

TABLE 8. Orthologous gene comparison and average nucleotide identity of Bacillus ndiopicus strain FF3^T with other compared genomes

	Bacillus ndiopicus	*Bacillus bataviensis*	*Bacillus coagulans* 2-6	*Bacillus coagulans* 36D1	*Bacillus isronensis*	*Lysinibacillus sphaericus*
Bacillus ndiopicus	3915	63.67	61.87	61.79	70.95	70.78
Bacillus bataviensis	1623	5207	64.71	64.49	63.31	63.61
Bacillus coagulans 2-6	1281	1617	2971	95.94	62.21	61.78
Bacillus coagulans 36D1	1359	1737	1824	3289	62.11	61.76
Bacillus isronensis	1934	1681	1332	1434	3883	69.18
Lysinibacillus sphaericus	1981	1669	1321	1413	1965	4584

TABLE 9. Antibiotic resistance genes in *Bacillus ndiopicus* strain FF3 genome

Gene	Size (aa)	Function	E-value	Antibiotic	GenBank ID
baca	275	Undecaprenyl pyrophosphate phosphatase	3e-66	Bacitracin	NC_009832
lmrb	465	ABC transporter system, macrolide-lincosamide-streptogramin B efflux pump	1e-128	Lincomycin	AB000617
vanA	266	D-Alanyl-D-alanine carboxypeptidase	1e-65	Vancomycin	AM410096
vatb	168	Virginiamycin A acetyltransferase	6e-13	Streptogramin A	U19459
str	282	Streptomycin resistance protein	2e-94	Streptomycin	P12055
bmr	390	Major facilitator superfamily transporter; multidrug resistance efflux pump	1e-128	Chloramphenicol/fluoroquinolone	D84432

The G+C content of the genome is 37.03%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers HG315675 and CCAP000000000, respectively. The type strain FF3^T (= CSUR P3025 = DSM 27837) was isolated from the skin of a healthy 39-year-old Senegalese man living in Ndiop, Senegal.

Acknowledgements

We thank C. Couderc for help in performing the MALDI-TOF analysis. We also thank F. Di Pinto for creating the electron microscope photos. This study was funded by the Méditerranée Infection Foundation.

Conflict of Interest

None declared.

References

[1] Cohn F. Untersuchungen über Bakterien. Beitrage zur Biologie der Pflanzen Heft 1872;1:127–224.
[2] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 2014;42(Database issue):D613–6.
[3] Castagnola E, Fioredda F, Barretta MA, et al. *Bacillus sphaericus* bacteraemia in children with cancer: case reports and literature review. J Hosp Infect 2001;48:142–5.
[4] Keita MB, Diene SM, Robert C, Raoult D, Fournier PE, Bittar F. Non-contiguous finished genome sequence and description of *Bacillus massiliogorillae* sp. nov. Stand Genomic Sci 2013;9:93–105.
[5] Bottone EJ. Bacillus cereus, a volatile human pathogen. Clin Microbiol Rev 2010;23:382–98.
[6] Mandell GL, Bennett JE, Dolin R. Principles and practice of infectious diseases. Amsterdam: Elsevier; 2010.
[7] Zhang YZ, Chen WF, Li M, et al. *Bacillus endoradicis* sp. nov., an endophytic bacterium isolated from soybean root. Int J Syst Evol Microbiol 2012;62:359–63.
[8] Pagani I, Liolios K, Jansson J, et al. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 2012;40:D571–9.
[9] Ramasamy D, Mishra AK, Lagier JC, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.
[10] Sentausa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect 2013;19:790–5.
[11] Lagier JC, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:1185–93.
[12] Trape JF, Tall A, Diagne N, et al. Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. Lancet Infect Dis 2011;11:925–32.
[13] Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 2013;195:413–8.
[14] Seng P, Drancourt M, Gouriet F, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.
[15] Field D, Garrity G, Gray T, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26:541–7.
[16] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 2010;11:119.

[17] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:48–53.
[18] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[19] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.
[20] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[21] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.
[22] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[23] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[24] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39:347–52.
[25] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinform 2011;12:124.
[26] Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 2014;64:352–6.
[27] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria*, and *Eukarya*. Proc Natl Acad Sci U S A 1990;87:4576–9.
[28] Skerman VBD, Sneath PHA. Approved list of bacterial names. Int J Syst Bacteriol 1980;30:225–420.
[29] Garrity GM, Holt J. The road map to the manual. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's manual of systematic bacteriology. 2nd ed., vol. 1. New York: Springer; 2001. p. 119–69.
[30] List of new names and new combinations previously effectively, but not validly, published. List no. 132. Int J Syst Evol Microbiol 2010;60:469–72.
[31] Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, et al., editors. Bergey's manual of systematic bacteriology. 2nd ed., vol. 3. New York: Springer-Verlag; 2009. p. 19–20.
[32] Prevot AR. In: Hauduroy P, Ehringer G, Guillot G, et al., editors. Dictionnaire des bactéries pathogènes. Paris: Masson; 1953. p. 692. Type genus: Bacillus Cohn 1872.
[33] Fischer A. Untersuchungen über bakterien. Jahrbücher für Wissenschaftliche Botanik 1895;27:1–163.
[34] Gibson T, Gordon RE. Genus I. *Bacillus* Cohn 1872, 174; Nom. gen. cons. Nomencl. Comm. Intern. Soc. Microbiol 1937, 28; Opin. A. Jud. Comm 1955, 39. In: Buchanan RE, Gibbons NE, editors. Bergey's manual of determinative bacteriology. 8th ed. Baltimore: Williams & Wilkins; 1974. p. 529–50.
[35] Mathews WC, Caperna J, Toerner JG, Barber RE, Morgenstern H. Neutropenia is a risk factor for Gram-negative *Bacillus* bacteremia in human immunodeficiency virus–infected patients: results of a nested case–control study. Am J Epidemiol 1998;148:1175–83.
[36] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
[37] Jee-Min L, Che OJ, Jung RL, Dong-Jin P, Chang-Jin K. *Bacillus* kribbensis sp. nov., isolated from a soil sample in Jeju, Korea. Int J Syst Evol Microbiol 2007;57:2912–6.
[38] Glazunova OO, Raoult D, Roux V. *Bacillus massiliensis* sp. nov., isolated from cerebrospinal fluid. Int J Syst Evol Microbiol 2006;56:1485–8.
[39] Jeroen H, Bram V, Niall AL, et al. *Bacillus novalis* sp. nov., *Bacillus vireti* sp. nov., *Bacillus soli* sp. nov., *Bacillus bataviensis* sp. nov. and *Bacillus drentensis* sp. nov., from the Drentse A grasslands. Int J Syst Evol Microbiol 2004;54:47–57.
[40] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7.
[41] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 2013;30:2725–9.

Article 10:

«Non-contiguous finished genome sequence and description of Haemophilus massiliensis sp. nov.»

Cheikh Ibrahima Lo[1,2*], Senthil Alias Sankar[1*], Bécaye Fall[3], Bissoume Sambe-Ba[3], Silman Diawara[3], Mamadou Wague Gueye[3], Oleg Mediannikov[1,2], Caroline Blanc-Tailleur[1], Boubacar Wade[3], Didier Raoult[1,2,4], Pierre-Edouard Fournier[1] and Florence Fenollar[1,2]*

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de Médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France Tel: +33 491 385 517

[2] Campus international UCAD-IRD, Dakar, Senegal

[3] Hôpital Principal, Dakar, Senegal

[4]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Soumis dans

Standards in Genomic Sciences

Résumé de l'article 10: ***«Non-contiguous finished genome sequence and description of Haemophilus massiliensis sp. nov. »***

La souche FF7^T a été isolée du liquide péritonéal d'une patiente Sénégalaise de 44 ans souffrant de péritonite pelvienne à l'Hôpital Principal de Dakar au Sénégal. L'analyse de la souche FF7^T par spectrométrie de masse MALDI-TOF montrant des scores de 1.32 à 1.56, n'a pas permis d'identification. La souche FF7^T présentait une similarité de séquence de l'ARN 16S ribosomique de 94,87% avec *Haemophilus parasuis,* l'espèce officielle la plus proche phylogénétiquement.

Nous avons réalisé des analyses phénotypiques et génomiques combinées concernant la souche FF7^T. Celle-ci est une bactérie à Gram négative, aéro-anaérobie, en forme de bâtonnet, et membre de la famille des *Pasteurellaceae*. Elle est immobile. La bactérie a une longueur moyenne de 2,6 µm (entre 2,0-3,2 µm) et un diamètre moyen de 0,35 µm (entre 0,2-0,5 µm). Sur gélose Columbia enrichi avec 5% de sang de mouton, les colonies ont une taille qui varie entre 0,5 à 1 mm et ne sont pas hémolytiques. La catalase et l'oxydase sont toutes deux positives.

De positives réactions ont été aussi observées avec les substrats suivants: l'acide phosphatase, la leucine arylamidase, l'estérase, la phosphatase alcaline et le Naphthol-AS-BI-phosphohydrolase, la L-arginine, l'esculine citrate de fer, et l'uréase. La souche FF7^T est sensible à la pénicilline, l'amoxicilline, l'amoxicilline-acide clavulanique, l'imipénème, la gentamicine, la ceftriaxone, et la

doxycycline mais résistante à la vancomycine, la nitrofurantoïne, et au triméthoprime-sulfaméthoxazole.

La souche FF7^T a un génome de 2 442 548 paires de base (un chromosome mais pas de plasmide), possède un taux de G+C de 46,0% et contient 2 319 protéines codantes et 67 gènes ARN, incluant 6 opérons ARNr.

Description d'*Haemophilus massiliensis* sp. nov.

Haemophilus massiliensis (mas.il.i.en'sis. L. gen. masc. n. massiliensis, de Massilia, le nom latin de la ville de Marseille, France, où la souche FF7^T a été caractérisée).

Les séquences de l'ARN 16S ribosomique et du génome de la souche FF7^T d'*Haemophilus massiliensis* (= CSUR P859 = DSM 28247) ont été déposées à GenBank sous les numéros d'accession respectifs : HG931334 et CCFL00000000. La souche type FF7^T a été isolée du liquide péritonéal d'une patiente Sénégalaise de 44 ans souffrant de péritonite pelvienne à Dakar au Sénégal.

Non-contiguous finished genome sequence and description of *Haemophilus massiliensis* sp. nov.

Cheikh Ibrahima Lo[1,2*], Senthil Alias Sankar[1*], Bécaye Fall[3], Bissoume Sambe-Ba[3], Silman Diawara[3], Mamadou Wague Gueye[3], Oleg Mediannikov[1,2], Caroline Blanc-Tailleur[1], Boubacar Wade[3], Didier Raoult[1,2,4], Pierre-Edouard Fournier[1] and Florence Fenollar[1,2]*

*Both authors contributed equally to this work

[1] Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
Tel: +33 491 385 517, Fax: +33 491 387 772

[2] Campus international UCAD-IRD, Dakar, Senegal

[3] Hôpital Principal, Dakar, Senegal

[4] Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

*Corresponding author: Florence Fenollar
(florence.fenollar@univ-amu.fr)

Abstract
Strain FF7^T was isolated from the peritoneal fluid of a 44-year-old woman who suffered from pelvic peritonitis. This strain exhibited a 16S rRNA sequence similarity of 94.87% 16S rRNA sequence identity with *Haemophilus parasuis,* the phylogenetically closest species with standing in nomenclature and a poor MALDI-TOF MS score (1.32 to 1.56) that does not allow any identification. Using a polyphasic study made of phenotypic and genomic analyses, strain FF7^T was a Gram-negative, areo-anaerobic rod and member of the family *Pasteurellaceae*. It exhibited a genome of 2,442,548 bp long genome (one chromosome but no plasmid) contains 2,319 protein-coding and 67 RNA genes, including 6 rRNA operons. On the basis of these data, we propose the creation of *Haemophilus massiliensis* sp. nov.

Keywords: *Haemophilus massiliensis*, genome, taxono-genomics, culturomics

Abbreviations
CSUR: Collection de Souches de l'Unité des Rickettsies
DSM: Deutsche Sammlung von Mikroorganismen
CNERS: National Ethics Committee of Senegal
MALDI-TOF MS: Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry
TE buffer: Tris-EDTA buffer
SDS: sodium dodecyl sulfate
MAGi: Marseille Average genomic identity
AGIOS: Average Genomic Identity of Orthologous Gene Sequences
GGDC: genome-to-genome distance calculator
dDDH: digital DNA-DNA hybridization

Introduction

The genus *Haemophilus* (Winslow *et al.* 1986) was described in 1986 [1] and currently includes 23 species. Some species of this genus are recognized human pathogens and cause meningitis, bacteremia, sinusitis, and/or pneumonia [2].

The current taxonomic classification of prokaryotes relies on a combination of phenotypic and genotypic characteristics [3; 4]; including 16S rRNA sequence similarity, G+C content and DNA-DNA hybridization. However, these tools suffer from various drawbacks, mainly due to their threshold values that are not applicable to all species or genera [5; 6]. With the development of cost-effective, high-throughput sequencing techniques, dozens of thousands of bacterial genome sequences have been made available in public databases [7]. Recently, we developed a strategy named taxono-genomics in which genomic and phenotypic characteristics, notably the MALDI-TOF-MS spectrum, are systematically compared to the phylogenetically-closest species with standing in nomenclature [8; 9].

The strain FF7^T (= CSUR P859 = DSM 28247) is the type strain of *Haemophilus massiliensis* sp. nov. It was isolated from the peritoneal fluid of a Senegalese woman suffering from pelvic peritonitis complicating a ruptured ovarian abscess. She was admitted to Hôpital Principal in Dakar, Senegal. *Haemophilus massiliensis* is a Gram-negative, aero-anaerobic, oxidase and catalase-positive and non-motile rod shaped bacterium. This microorganism was cultivated as part of the MALDI-TOF-MS implementation in Hôpital Principal in Dakar, aiming at improving the routine laboratory identification of bacterial strains in Senegal [10]. Here, we present a summary classification and a set of features for *Haemophilus massiliensis* sp. nov. together with the description of the complete genome sequencing and annotation. These characteristics support the circumscription of the species *Haemophilus massiliensis*.

Organism information

Classification and features

Since July 2012, the Hôpital Principal in Dakar, Senegal, has been equipped with a MALDI-TOF MS (VITEK® MS RUO, bioMérieux, Marcy l'Etoile, France) to improve the microbiology laboratory workflow by enabling rapid bacterial strain identification. Isolates that are poorly identified using MALDI-TOF-MS are referred to the URMITE laboratory in Marseille, France, for further characterization. In June 2013, a bacterial strain (Table 1) was isolated by cultivation on 5% sheep blood-enriched Columbia agar (BioMérieux, Marcy l'Etoile, France) of a peritoneal fluid specimen obtained from a 44-year-old Senegalese woman who suffered from pelvic peritonitis that had complicated a ruptured ovarian abscess [10]. The strain could not be identified using MALDI-TOF-MS. Strain FF7^T exhibited a 94.87% 16S rRNA sequence identity with *Haemophilus parasuis* (GenBank accession number AY362909), the phylogenetically-closest bacterial species with a validly published name (Figure 1). These values were lower than the 98.7% 16S rRNA gene sequence threshold recommended by Meier-Kolthoff et al., 2013 to delineate a new species within phylum *Proteobacteria* without carrying out DNA-DNA hybridization [20].

Different growth temperatures (25°C, 30°C, 37°C, 45°C and 56°C) were tested. Growth was obtained between 25 and 45°C, with the optimal growth temperature being 37°C. Colonies were 0.5 mm in diameter and non-haemolytic on 5% sheep blood-enriched Columbia agar (BioMérieux). Gram staining showed rod-shaped Gram-negative bacilli that were not motile and unable to form spores (Figure 2). In electron microscopy, cells had a mean length of 2.6 µm (range 2.0-3.2 µm) and width of 0.35 µm (range 0.2-0.5 µm) (Figure 3). Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux), and under aerobic conditions, with or without 5% CO_2. Optimal growth was observed at 37°C under aerobic and microaerophilic

conditions.
Strain FF7^T exhibited oxidase and catalase activities. Using an API ZYM strip (BioMérieux), positive reactions were observed for acid phosphatase, leucine arylamidase, esterase, alkaline phosphatase and Naphthol-AS-BI-phosphohydrolase. Negative reactions were noted for α-chymotrypsin, cystine arylamidase, valine arylamidase, trypsin, α-glucosidase, β-glucosidase, esterase-lipase, leucine arylamidase, α-galactosidase, β-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase and N-acetyl-β-glucosaminidase. Using API 20NE (BioMérioeux), positive reactions were obtained for L-arginine, esculin, ferric citrate and urea but negative reactions were observed for D-glucose, L-arabinose, D-maltose, D-mannose, D-mannitol, potassium gluconate and N-acetyl-glucosamine. *Haemophilus massiliensis* strain FF7^T is susceptible to penicillin, amoxicillin, amoxicillin/clavulanic acid, imipenem, gentamicin, ceftriaxone and doxycycline but resistant to vancomycin, nitrofurantoin and trimethoprim/sulfamethoxazole. Five species validly published names in the *Haemophilus* genus were selected to make a phenotypic comparison with our new species named *Haemophilus massiliensis* detailed in Table 2.

Extended features descriptions

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was performed using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany), as reported elsewhere [28]. The scores previously established by Bruker Daltonics allowing validating or not the identification of species compared to the database of the instrument were applied. Briefly, a score ≥ 2 with a species with a validly published name provided allows the identification at the species level; a score ≥ 1.7 and < 2 allows the identification at the genus level; and a score < 1.7 does not allow any identification. We performed 12 distinct deposits from 12 isolated colonies of strain FF4^T. Two microliters of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 2.5% trifluoroacetic-acid were distributed on each smear and submitted at air drying for five minutes. For strain FF7^T, scores ranging from 1.32 to 1.56 were obtained. Therefore the isolate could not be classified within any known species. The reference mass spectrum from strain FF7^T was incremented in our database (Figure 4). Finally, the gel view showed the spectral differences with other members of the genus *Haemophilus* (Figure 5).

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its 16S rRNA similarity, phylogenetic position and phenotypic differences with other members of the genus *Haemophilus*, and is part of a study aiming at using MALDI-TOF-MS for the routine identification of bacterial isolates in Hôpital Principal in Dakar [10]. It is the eleventh genome of a *Haemophilus* species and the first genome of *Haemophilus massiliensis* sp. nov. A summary of the project information is shown in Table 3. The Genbank accession number is CCFL00000000 and consists of 148 contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [28].

Growth conditions and genomic DNA preparation

Haemophilus massiliensis sp. nov., strain FF7^T (= CSUR P859= DSM 28247) was grown aerobically on 5 % sheep blood-enriched Columbia agar (BioMérieux) at 37°C. Bacteria grown on four Petri dishes were resuspended in 5x100 µL of TE buffer; 150 µL of this suspension was diluted in 350 µL TE buffer 10X, 25 µL proteinase K and 50 µL sodium dodecyl sulfate (SDS) for lysis treatment. This preparation was incubated overnight at 56°C. Extracted DNA was purified using 3 successive phenol-chloroform extractions and ethanol

precipitations. Following centrifugation, the DNA was suspended in 65 µL EB buffer. The genomic DNA (gDAN) concentration was measured at 14.7ng/µl using the Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, USA).

Genome sequencing and assembly

Genomic DNA of *H. massiliensis* FF7^T was sequenced on the MiSeq sequencer (Illumina, San Diego, CA, USA) with the Mate-Pair strategy. The gDNA was barcoded in order to be mixed with 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina). The Mate-Pair library was prepared with 1µg of genomic DNA using the Nextera Mate-Pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a Mate-Pair junction adapter. The pattern of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1 kb up to 10 kb with an optimal size at 4.08 kb. No size selection was performed and only 464 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 569 bp on the Covaris S2 device in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies) and the final library concentration was measured at 24.42 nmol/L. The libraries were normalized at 2nM and pooled. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 39-hour-run in a 2x251-bp. Total information of 10.1Gb was obtained from a 1,189 K/mm2 cluster density with a cluster passing quality control filters of 99.1% (22,579,000 clusters). Within this run, the index representation for *Haemophilus massiliensis* was 9.72%. The 1,976,771 paired reads were filtered according to the read qualities. These reads were trimmed, then assembled using the CLC genomicsWB4 software. Finally, the draft genome of *Haemophilus massiliensis* consists of 9 scaffolds with 148 contigs and generated a genome size of 2.4Mb with a 46.0% G+C content.

Genome annotation

Open Reading Frames (ORFs) were predicted using Prodigal [29] with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank database [30] and the Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAScanSE tool [31] was used to find tRNA genes, whereas ribosomal RNAs were found using RNAmmer [32] and BLASTn against the GenBank database. Lipoprotein signal peptides and the number of transmembrane helices were predicted using SignalP [33] and TMHMM [34] respectively. ORFans were identified if their BLASTP E-value was lower than 1e-03 for alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E-value of 1e-05. Such parameter thresholds have already been used in previous works to define ORFans. Artemis [35] was used for data management and DNA Plotter [36] for visualization of genomic features. The Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [37]. To estimate the mean level of nucleotide sequence similarity at the genome level, we used the average genomic identity of gene sequences (AGIOS) home-made software [9]. Briefly, this software combines the Proteinortho software [38] for detecting orthologous proteins in pairwise genomic comparisons, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm. The script created to calculate AGIOS values was named MAGi

(Marseille Average genomic identity) and is written in perl and bioperl modules. GGDC analysis was also performed using the GGDC web server (http://ggdc.dsmz.de) as previously reported [40, 41].

Genome properties

The genome of *Haemophilus massiliensis* strain FF7^T is 2,442,548 bp-long with a 46.0% G+C content (Table 4, Figures 6 and 7). Of the 2,386 predicted genes, 2,319 were protein-coding genes and 67 were RNA genes, including six complete rRNA operons. A total of 1,885 genes (79.50%) were assigned a putative function. A total of 36 genes were identified as ORFans (1.51%). The remaining genes were annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Tables 4. The distribution of genes into COGs functional categories is presented in Table 5.

Insights from the genome sequence

Extended insights

Here, we compared the genome sequences of *Haemophilus massiliensis* strain FF7^T (GenBank accession number CCFL00000000) with those of *Haemophilus parasuis* strain SH0165 (CP001321), *Haemophilus influenzae* strain Rd KW20 (L42023), *Aggregatibacter segnis* strain ATCC 33393^T (AEPS00000000), *Haemophilus sputorum* strain CCUG 13788^T (AFNK00000000), *Haemophilus pittmaniae* strain HK 85 (AFUV00000000), *Haemophilus aegyptius* strain ATCC 1111^T (AFBC00000000), *Haemophilus parainfluenzae* strain ATCC 33392^T (AEWU00000000), *Haemophilus haemolyticus* strain M21621 (AFQQ00000000), *Haemophilus ducreyi* strain 35000HP (AE017143) and *Haemophilus parahaemolyticus* strain HK385 (AJSW00000000).

The draft genome of *Haemophilus massiliensis* has a larger size than that of *H. parasuis*, *H. influenzae, A. segnis* , *H. sputorum*, *H. pittmaniae*, *H. aegyptius*, *H. parainfluenzae*, *H. haemolyticus*, *H. ducreyi* and *H. parahaemolyticus* (2.44, 2.27, 1.83, 1.99, 2.14, 2.18, 1.92, 2.11, 2.09, 1.7 and 2.03 Mb, respectively). The G+C content of *Haemophilus massiliensis* is higher than those of *H. parasuis*, *H. influenzae, A. segnis*, *H. sputorum*, *H. pittmaniae*, *H. aegyptius*, *H. parainfluenzae*, *H. haemolyticus, H. ducreyi* and *H. parahaemolyticus* (46.0, 40.0, 38.2, 42.5, 39.7, 42.5, 38.1, 39.1, 38.4, 38.2 and 40.1%, respectively). As it has been suggested in the literature that the G+C content deviation is at most 1% within species, these data are an additional argument for the creation of a new taxon [41].

The gene content of *Haemophilus massiliensis* is larger than those of *H. parasuis, H. influenzae, A. segnis*, *H. sputorum*, *H. aegyptius, H. parainfluenzae*, *H. haemolyticus*, *H. ducreyi* and *H. parahaemolyticus* (2,319, 2,299, 1,765, 1,956, 2,072, 2,020, 2,068, 2,056, 1,717 and 1,980, respectively) but smaller than that of *H. pittmaniae* (2,390). However the distribution of genes into COG categories was similar in all compared genomes (Figure 7). In addition, *Haemophilus massiliensis* shared 2,021, 1,956, 2,020, 1,717, 1,977, 1,610, 1,980, 2,010, 2,390 and 2,123 orthologous genes with *H. parasuis*, *A. segnis*, *H. aegyptius*, *H. ducreyi, H. haemolyticus*, *H. influenzae*, *H. parahaemolyticus, H. parainfluenzae*, *H. pittmaniae* and *H. sputorum*, respectively (Table 6). Among species with standing in nomenclature, AGIOS values ranged from 71.19 between *H. pittmaniae* and *H. ducreyi* to 97.31% between *H. influenzae* and *H. aegyptius*. When compared to other species, *Haemophilus massiliensis exhibited* AGIOS values ranging from 70.00 with *H. ducreyi* to 74.19 with *A. segnis*. We obtained similar results using the GGDC software, as dDDH values ranged from 0.201 to 0.777 between studied species, and were 0.248 between *Haemophilus massiliensis and Haemophilus parasuis*. These values confirm the status of *Haemophilus massiliensis* as a new species.

Conclusions

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Haemophilus massiliensis* sp. nov. that contains strain FF7^T. The strain was isolated from a peritoneal fluid specimen from a 44-year-old Senegalese woman admitted to Hôpital Principal in Dakar, Senegal.

Taxonomic and nomenclatural proposals

Description of *Haemophilus massiliensis* strain FF7^T sp. nov.

Haemophilus massiliensis (mas.il.i.en'sis. L. gen. masc. n. massiliensis, of Massilia, the Latin name of Marseille where strain FF7^T was characterized). On 5% sheep blood-enriched Columbia agar (BioMérieux), colonies are round and light with a size of 0.5-1 mm. Colonies are not haemolytic on blood agar. Cells are Gram-negative and not motile, with a mean length of 2.6 µm (range 2.0-3.2 µm) and a mean diameter of 0.35 µm (range 0.2-0.5 µm). Catalase and oxidase reactions are positive. Positive reactions are also observed with the following substrates (using API ZYM strip, BioMérieux): acid phosphatase, leucine arylamidase, esterase, alkaline phosphatase and Naphthol-AS-BI-phosphohydrolase. Negative reactions are obtained for α-chymotrypsin, cystine arylamidase, valine arylamidase, trypsin, α-glucosidase, β- glucosidase, esterase-lipase, leucine arylamidase, α-galactosidase, β-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase and N-acetyl-β-glucosaminidase. Using API 20NE strip (BioMérieux), positive reactions are obtained for L-arginine, esculin, ferric citrate and urea but negative reactions are observed for D-glucose, L-arabinose, D-maltose, D-mannose, D-mannitol, potassium gluconate and N-acetyl-glucosamine. *Haemophilus massiliensis* strain FF7^T is susceptible to penicillin, amoxicillin, amoxicillin/clavulanic acid, imipenem, gentamicin, ceftriaxone and doxycycline but resistant to vancomycin, nitrofurantoin and trimethoprim/sulfamethoxazole. The G+C content of the genome is 46.0%. The 16S rRNA and genome sequences of *Haemophilus massiliensis* strain FF7^T (= CSUR P859 = DSM 28247) are deposited in GenBank under accession numbers HG931334 and CCFL00000000, respectively. The type strain FF7^T was isolated from the peritoneal fluid of a 44-year-old Senegalese woman suffering from pelvic peritonitis in Dakar, Senegal.

Authors' contributions

CIL performed the phenotypic characterization of the bacterium and drafted the manuscript. SAS performed the genomic analyses and drafted the manuscript. BF participated in the design of the study and helped to draft the manuscript. BSB, SD and MWG performed the phenotypic characterization of the bacterium and helped to draft the manuscript. OM participated in its design and helped to draft the manuscript. CBT performed the genomic sequencing and helped to draft the manuscript. BW conceived the study and helped to draft the manuscript. DR conceived the study and helped to draft the manuscript. PEF and FF conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the Xegen Company (www.xegen.fr) for automating the genomic annotation process. This study was funded by the Mediterranée-Infection Foundation.

References

1. Winslow CEA, Broadhurst J, Buchanan RE, Krumwiede C, Rogers LA, Smith GH. The Families and Genera of the Bacteria: Preliminary Report of the Committee of the Society of American Bacteriologists on Characterization and Classification of Bacterial Types. *J Bacteriol* 1917; 2:505-566.
2. Musher DM. *Haemophilus* species. In book: Medical Microbiology, Edition: 4th, Chapter 30, Publisher: University of Texas Medical Branch at Galveston, Editors: Samuel Baron. 1996.
3. Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 2010; 60:249-266.
4. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today* 2006; 33:152-155.
5. Wayne LG, Brenner DJ, Colwell PR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E and others. Report of the ad hoc committee on reconciliation of approaches to bacterial systematic. *Int J Syst Bacteriol* 1987; 37:463-464.
6. Rosselló-Móra R. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E (ed), Molecular Identification, Systematics, and population Structure of Prokaryotes. Springer, Berlin, 2006; p. 23-50.
7. Reddy TBK, Thomas A, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos E and Kyrpides N. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification. *Nucl Acids Res* 2014; 43: 1099-1106.
8. Lagier JC, El Karkouri K, Mishra AK, Robert C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of Enterobacter massiliensis sp. nov. *Stand Genomic Sci* 2013; 7:399-412. http://dx.doi: 10.4056/sigs.3396830.
9. Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi-Tamisier M, Sentausa E, Raoult D, Fournier PE. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. *Int J Syst Evol Microbiol* 2014; 64:384-391.
10. Fall B, Lo CI, Samb-Ba B, Perrot N, Diawara S, Gueye MW, Sow K, Aubadie-Ladrix M, Mediannikov O, Sokhna C, Diemé Y, Chatellier S, Wade B, Raoult D, Fenollar F. The ongoing revolution of MALDI-TOF mass spectrometry for microbiology reaches tropical Africa. *Am J Trop Med Hyg* 2015. In press.
11. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; 87:4576-4579.
12. Garrity GM, Bell JA, Lilburn T: Phylum XIV. *Proteobacteria* phyl. nov. Bergey's Manual of Systematic Bacteriology 2005, 2, Part B: 1.
13. Garrity GM, Bell JA, Lilburn T: Class III. *Gammaproteobacteria* class. nov. Bergey's Manual of Systematic Bacteriology 2005, 2, Part B: 1.
14. List Editor: Validation of publication of new names and new combinations previously effectively published outside the IJSEM. List no. 106. *Int J of Syst Evol Microbiol* 2005, 55: 2235-2238.
15. Garrity GM, Bell JA, Lilburn T: Order XIV. *Pasteurellales* ord. nov. Bergey's Manual of Systematic Bacteriology 2005, 2, Part B: 850.
16. Pohl S: Reklassifizierung der gattung *Actinobacillus* Brumpt 1910, Haemophilus Winslow et al. 1971 und Pasteurella Trevisan 1887 anhand phänotypischer und molekularer daten, insbesondere der DNS-verwandtschaften bei DNS: DNS-hybridisierung in vitro und vorschlag einer neuen familie, *Pasteurellaceae*. 1979.
17. List Editor: Validation List no. 7. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. *Int J Syst Bacteriol* 1981, 31: 382-383.

18. Skerman VBD, McGowan V, Sneath PHA: Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; 30: 225-420.
19. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25:25-29.
20. Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? *Arch Microbiol*. 2013; 195(6):413-8. http://dx.doi:10.1007/s00203-013-0888-4.
21. Nørskov-Lauritsen N, Bruun B, Andersen C, Kilian M: Identification of haemolytic *Haemophilus* species isolated from human clinical specimens and description of *Haemophilus sputorum* sp. nov. *Int J Med Microbiol* 2012; 302: 78-83.
22. List Editor: List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2012, 62: 1443-1445.
23. Nørskov-Lauritsen N, Bruun B, Andersen C, Kilian M. Multilocus sequence phylogenetic study of the genus *Haemophilus* with description of *Haemophilus pittmaniae* sp. nov. *Int J Syst Evol Microbiol* 2005; 55:449-456.
24. Inzana TJ, Johnson JL, Shell L, Møller K, Kilian M. Isolation and characterization of a newly identified *Haemophilus* species from cats: "*Haemophilus felis*". *Journal of clinical Microbiology* 1992; 30:2108-2112.
25. List Editor: Validation of publication of new names and new combinations previously effectively published outside the IJSB. List No. 69. *Int J Syst Bacteriol* 1999; 49: 341-342.
26. Biberstein EL, White DC. A proposal for the establishment of two new *Haemophilus* species. *Journal of Medical Microbiology* 1969; 2: 75-78.
27. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009; 49:543-551.
28. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol*. 2008; 26:541-547.
29. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010; 11:119.
30. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2012; 40:D48-D53.
31. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; 25:955-964.
32. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl Acids Res* 2007; 35:3100-3108.
33. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; 340:783-795.
34. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; 305:567-580.
35. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics* 2000; 16:944-945.
36. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009; 25:119-120.
37. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004; 14:1394-1403.
38. Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of

(Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011; 12:124.
39. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32:1792-1797.
40. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013; 14:60.
41. Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int J Syst Evol Microbiol* 2014; 64:352-356.
42. Auch AF, Von Jan M, Klenk HP, GökerM. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2010; 2:117-134. http://dx.doi.org/10.4056/sigs.531120.
43. Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* 2010; 2:142-8. http://dx.doi.org/10.4056/sigs.541628.

Table 1. Classification and general features of *Haemophilus massiliensis* strain FF7^T

MIGS ID	Property	Term	Evidence code[a]
	Current classification	Domain: *Bacteria*	TAS [11]
		Phylum: *Proteobacteria*	TAS [12]
		Class: *Gammaproteobacteria*	TAS [13; 14]
		Order: *Pasteurellales*	TAS [14;15]
		Family: *Pasteurellaceae*	TAS [16; 17]
		Genus: *Haemophilus*	TAS [1;18]
		Species: *Haemophilus massiliensis*	IDA
		Type strain: FF7^T	IDA
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Not motile	IDA
	Sporulation	Non-spore forming	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37°C	IDA
	pH range; Optimum	7.2-7.4; 7.3	
	Carbon source	Unknown	NAS
	Energy source	Unknown	NAS
MIGS-6	Habitat	Human peritoneal fluid	IDA
MIGS-6.3	Salinity	Unknown	
MIGS-22	Oxygen requirement	Aero-anerobic	IDA
MIGS-15	Biotic relationship	Free living	IDA
MIGS-14	Pathogenicity	Unknown	
	Biosafety level	2	
	Isolation	Human	IDA
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection time	June 2013	IDA
MIGS-4.1	Latitude	14.6937000	IDA
MIGS-4.1	Longitude	-17.4440600	IDA
MIGS-4.3	Depth	Surface	IDA
MIGS-4.4	Altitude	12 m above sea level	IDA

[a]Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from http://www.geneontology.org/GO.evidence.shtml of the Gene Ontology project [19]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

Table 2. Differential characteristics of *Haemophilus massiliensis* strain FF7^T (data from this study), *H. influenzae* strain ATCC 33391^T [1; 18], *H. sputorum* strain CCUG 13788^T [21;22], *H. pittmaniae* strain HK 85^T [23; 26], *Haemophilus felis* strain TI189^T [24;25] and *H. parasuis* strain ATCC 19417^T [18;26]. na = data not available.

Characters	*Haemophilus massiliensis*	*H. influenzae*	*Haemophilus felis*	*H. parasuis*	*H. sputorum*	*H. pittmaniae*
Cell diameter (μm)	0.2-0.5	0.3-0.5	0.4-0.5	na	0.3–0.5	na
Gram stain	-	-	-	-	-	-
Motility	-	-	na	-	-	-
Endospore formation	-	-	-	variable	na	na
Production of						
Alkaline phosphatase	+	+	+	+	+	+
Acid phosphatase	+	na	na	na	na	+
Catalase	+	+	+	+	+	+
Oxidase	+	+	-	-	+	+
β-Haemolysis	-	-	variable	-	+	+
Urease	-	+	-	-	+	-
Indole	-	+	-	-	-	-
Nitrate reductase	+	+	+	+	na	na
α-galactosidase	-	na	na	-	na	na
β-Galactosidase	-	na	-	na	+	+
α-glucosidase (PNPG)	+	-	+	variable	na	-
β-glucosidase	-	-	na	na	na	-
Esterase	+	na	na	na	na	na
Esterase lipase	-	na	na	na	na	na
N-acetyl-β-glucosaminidase	-	na	na	na	-	na
Utilization of						
D-fructose	+	-	na	+	+	+
D-mannose	-	-	+	+	-	+
D-Xylose	+	+	-	-	-	-
D-glucose	+	+	-	-	+	+
Habitat	Peritoneal fluid	Clinical samples	Cat	Respiratory tract of swine	Human tooth	Human saliva

434 **Table 3.** Project information

MIGS ID	Property	
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Mate-Pair 3.1kb library
MIGS-29	Sequencing platforms	Illumina Miseq
MIGS-31.2	Fold coverage	42.54
MIGS-30	Assemblers	CLC GENOMICSWB4
MIGS-32	Gene calling method	Prodigal
	Locus Tag	Not indicated
	Genbank ID	CCFL00000000
	Genbank Date of Release	August 22, 2014
	GOLD ID	Ga0059233
	Bioproject	PRJEB5521
MIGS-13	Source material identifier	DSM 28247
	Project relevance	MALDI-TOF-MS implementation in Dakar

435 **Table 4.** Genome statistics.

Attribute	Genome (total)	
	Value	% of total[a]
Genome size (bp)	2,442,548	100
DNA coding (bp)	2,181,795	89.35
DNA G+C (bp)	1,123,572	46.0
DNA scaffolds	10	-
Total genes	2,386	100
Protein coding genes	2,319	97.19
RNA genes	67	2.80
Pseudo genes	N/D[b]	-
Gens in internal clusters	N/D[b]	-
Genes with function prediction	1,885	79.00
Genes assigned to COGs	2,093	87.72
Genes with Pfam domains	1,419	59.47
Genes with signal peptides	188	7.87
Genes with transmembrane helices	445	18.65
ORFan genes	36	1.50
CRISPR repeats	2	0.08

436 [a] The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the
437 annotated genome.
438 [b] N/D = not determined.
439

Table 5. Number of genes associated with the 25 general COG functional categories

Code	Value	% of total[a]	Description
J	160	6.90	Translation
A	1	0.04	RNA processing and modification
K	150	6.47	Transcription
L	131	5.65	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	28	1.21	Cell cycle control, mitosis and meiosis
V	37	1.60	Defense mechanisms
T	46	1.98	Signal transduction mechanisms
M	135	5.82	Cell wall/membrane biogenesis
N	6	0.26	Cell motility
W	9	0.39	Extracellular structures
U	58	2.50	Intracellular trafficking and secretion
O	109	4.70	Posttranslational modification, protein turnover, chaperones
C	163	7.03	Energy production and conversion
G	228	9.83	Carbohydrate transport and metabolism
E	234	10.09	Amino acid transport and metabolism
F	63	2.72	Nucleotide transport and metabolism
H	117	5.05	Coenzyme transport and metabolism
I	62	2.67	Lipid transport and metabolism
P	152	6.55	Inorganic ion transport and metabolism
Q	34	1.47	Secondary metabolites biosynthesis, transport and catabolism
R	258	11.13	General function prediction only
S	174	7.50	Function unknown
-	226	9.53	Not in COGs

[a] The total is based on the total number of protein coding genes in the annotated genome.

Table 6. dDDH values and AGIOS values obtained (lower left). The numbers of proteins per genome are indicated by bold numbers. Digital DDH similarities between the genomes were calculated using GGDC web server version 2.0. under recommend setting [42; 43] ; formula 2 is recommended, particularly for draft genomes. [a]*Haemophilus massiliensis*, [b]*Haemophilus parasuis*, [c]*Haemophilus influenzae*, [d]*Haemophilus aegyptius*, [e]*Aggregatibacter segnis*, [f]*Haemophilus haemolyticus*, [g]*Haemophilus parainfluenzae*, [h]*Haemophilus ducreyi*, [i]*Haemophilus sputorum*, [j]*Haemophilus pittmaniae* and [k]*Haemophilus parahaemolyticus*.

	HM[a]	HPA[b]	HI[c]	HA[d]	AE[e]	HH[f]	HP[g]	HD[h]	HS[i]	HPT[j]	HPH[k]
HM	**2,319**	0.248	0.222	0.203	0.204	0.202	0.243	0.285	0.232	0.235	0.201
HPA	70.64	**2,021**	0.296	0.292	0.244	0.278	0.236	0.237	0.262	0.283	0.251
HI	72.76	72.75	**1,610**	0.777	0.234	0.433	0.259	0.252	0.279	0.252	0.239
HA	72.80	72.67	97.31	**2,020**	0.231	0.434	0.237	0.244	0.271	0.242	0.235
AE	74.19	71.93	75.72	75.69	**1,956**	0.235	0.243	0.267	0.255	0.246	0.232
HH	72.72	72.64	91.85	91.80	75.72	**1,977**	0.240	0.250	0.284	0.247	0.246
HP	70.16	75.40	74.07	73.98	71.97	73.79	**1,980**	0.239	0.218	0.293	0.227
HD	70.00	74.81	72.47	72.36	71.34	72.33	75.34	**1,717**	0.228	0.270	0.251
HS	70.23	75.06	73.46	73.39	72.33	73.54	78.00	75.68	**2,123**	0.319	0.280
HPT	72.55	71.51	76.38	76.48	74.95	76.64	71.88	71.19	72.88	**2,390**	0.269
HPH	72.67	72.71	79.69	79.70	76.20	79.96	73.36	72.30	74.14	78.94	**2,010**

Table S1. Associated MIGS record.

MIGS-ID	field name	Description
MIGS-1	Submit to INSDC/Trace archives	Not reported
1.1	PID	Not reported
1.2	Trace Archive	Not reported
MIGS-2	MIGS CHECK LIST TYPE	Not reported
MIGS-3	Project Name	Implementation of MALDI-TOF MS
MIGS-4	Geographic Location	Dakar, France
4.1	Latitude	14.6937000
4.2	Longitude	-17.4440600
4.3	Depth	Surface
4.4	Altitude	12 m
MIGS-5	Time of Sample collection	June 2013
MIGS-6	Habitat (EnvO)	Pus from ovarian abscess
6.1	Temperature	37°C
6.2	pH	Not reported
6.3	Salinity	Not reported
6.4	Chlorophyll	Not reported
6.5	Conductivity	Not reported
6.6	light intensity	Not reported
6.7	dissolved organic carbon (DOC)	Not reported
6.8	Current	Not reported
6.9	atmospheric data	Not reported
6.10	Density	Not reported
6.11	Alkalinity	Not reported
6.12	dissolved oxygen	Not reported
6.13	particulate organic carbon (POC)	Not reported
6.14	Phosphate	Not reported
6.15	Nitrate	Not reported
6.16	Sulfates	Not reported
6.17	Sulfides	Not reported
6.18	primary production	Not reported
MIGS-7	Subspecific genetic lineage	Not reported
MIGS-9	Number of replicons	1
MIGS-10	Extrachromosomal elements	0
MIGS-11	Estimated Size	2,442,548 bp
MIGS-12	Reference for biomaterial or Genome report	Not reported
MIGS-13	Source material identifiers	Not reported
MIGS-14	Known Pathogenicity	Not reported
MIGS-15	Biotic Relationship	Not reported
MIGS-16	Specific Host	Not reported
MIGS-17	Host specificity or range (taxid)	Not reported
MIGS-18	Health status of Host	Sick
MIGS-19	Trophic Level	Not reported
MIGS-22	Relationship to Oxygen	Aerobic
MIGS-23	Isolation and Growth conditions	blood-enriched Columbia agar at 37°C

MIGS-27	Nucleic acid preparation	Phenol-chloroform extraction
MIGS-28	Library construction	Illumina Miseq
28.1	Library size	3.1kb
28.2	Number of reads	3953542
28.3	Vector	Not reported
MIGS-29	Sequencing method	Illumina Miseq
MIGS-30	Assembly	CLCGENOMICSWB4
30.1	Assembly method	Not reported
30.2	estimated error rate	Not reported
30.3	method of calculation	Not reported
MIGS-31	Finishing strategy	Not reported
31.1	Status	Draft genome, unfinished
31.2	Coverage	42.54X
31.3	Contigs	148
MIGS-32	Relevant SOPs	Not reported
MIGS-33	Relevant e-resources	Not reported

Figure legends.

Figure 1. Phylogenetic tree showing the position of *Haemophilus massiliensis* strain FF7^T relative to other type strains within the *Pasteurellaceae*. The strains (type = ^{T}) and their corresponding GenBank accession numbers for 16S rRNA genes (in parenthesis are indicated genome GenBank numbers, when available) are: *Haemophilus haemoglobinophilus* strain ATCC 19416^T, AY362907; *Avibacterium paragallinarum* strain NCTC 11296^T, AY498868 (GA: CBMK000000000); *Haemophilus felis* strain TI189^T, AF224292; *H. massilliensis* strain FF7^T, HG931334 (GA: CCFL00000000); *H. parasuis* strain ATCC 19417^T, AY362909; *H. influenzae* strain ATCC 33391^T, M35019; *H. aegyptius* strain ATCC 11116^T, AY362905 (GA: AFBC00000000); *Aggregatibacter segnis* strain ATCC 33393^T, M75043 (GA: AEPS00000000); *H. haemolyticus* strain ATCC 33390^T, JN175335 (GA: AFQO00000000); *H. parahaemolyticus* strain ATCC 10014^T, AJ295746 (GA: AJSW00000000); *H. ducreyi* strain CIP 54.2, M63900; *H. sputorum* strain CCUG 13788^T, JF506642 (GA: AFNK00000000); *H. pittmaniae* strain HK 85^T, AJ290755 (GA: AFUV00000000); *H. parainfluenzae* strain ATCC 33392^T, AY362908 (GA: AEWU00000000); *Escherichia coli* strain ATCC 11775^T, X80725 (GA: BAIM00000000). Sequences were aligned using MUSCLE [39], and phylogenetic tree inferred using the Maximum Likelihood method with Kimura 2-parameter model from MEGA software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. The scale bar represents a rate of substitution per site of 0.01. *Escherichia coli* strain ATCC 11775^T was used as outgroup.

Figure 2. Gram staining of *Haemophilus massiliensis* sp. nov. strain FF7^T.

Figure 3. Transmission electron microscopy of *Haemophilus massiliensis* strain FF7^T. Cells are observed on a Tecnai G20 transmission electron microscope operated at 200 keV. The scale bar represents 500 nm.

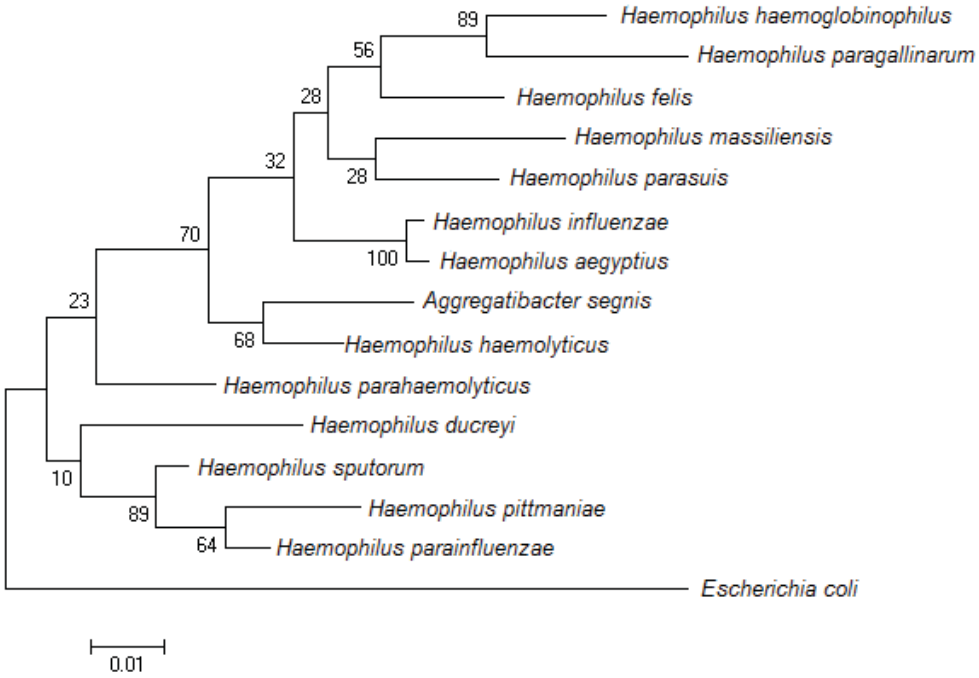
Figure 4. Reference mass spectrum from *Haemophilus massiliensis* strain FF7^T. Spectra from 12 individual colonies were compared and a reference spectrum was generated.

Figure 5. Gel view comparing *Haemophilus massiliensis* strain FF7^T to the members of the family *Pasteurellaceae.* The gel view displays the raw spectra of all loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Gray scale scheme code. The color bar and the right y-axis indicate the relation between the color a peak is displayed with and the peak intensity in arbitrary units. Displayed species are indicated on the left.

Figure 6. Graphical circular map of the *Haemophilus massiliensis* strain FF7^T chromosome. From the outside in, the outer two circles show open reading frames oriented in the forward (colored by COG categories) and reverse (colored by COG categories) directions, respectively. The third circle marks the tRNA genes (green). The fourth circle shows the G+C% content plot. The inner-most circle shows GC skew, purple indicating negative values whereas olive for positive values.

Figure 7. Distribution of functional classes of predicted genes in the genomes from *H. massiliensis* (HM), *H. parasuis* (HPA), *A. segnis* (AE), *H. aegyptius* (HA), *H. ducreyi* (HD), *H. haemolyticus* (HH), *H. influenzae* (HI), *H. parahaemolyticus* (HP), *H. parainfluenzae* (HPI), *H. pittmaniae* (HPT) and *H. sputorum* (HS) chromosomes according to the clusters of orthologous groups of proteins.

494 **Figure 1.**
495



496
497

Figure 2.

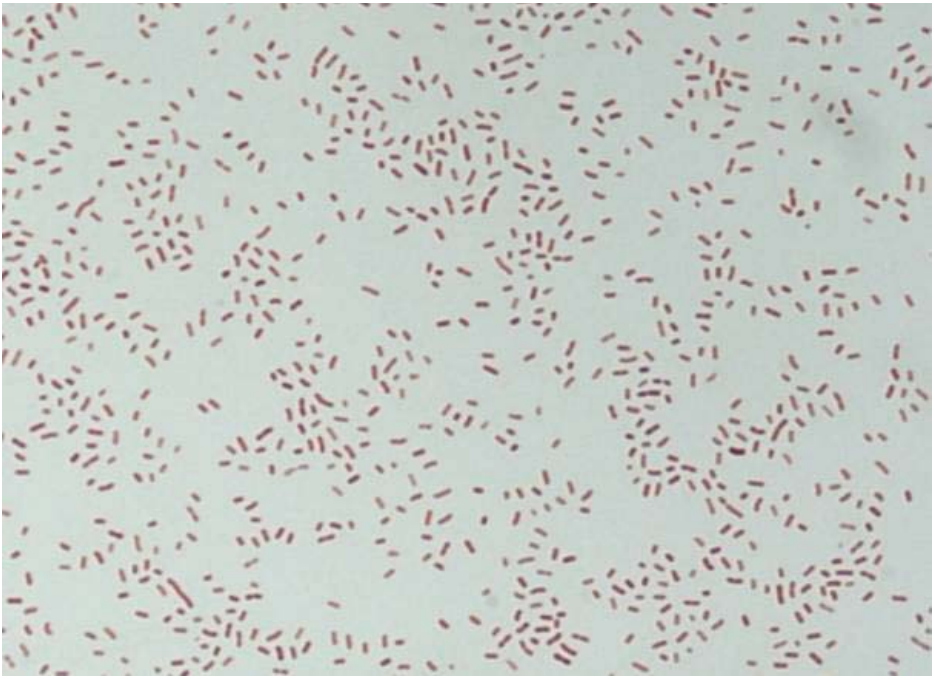


Figure 3.

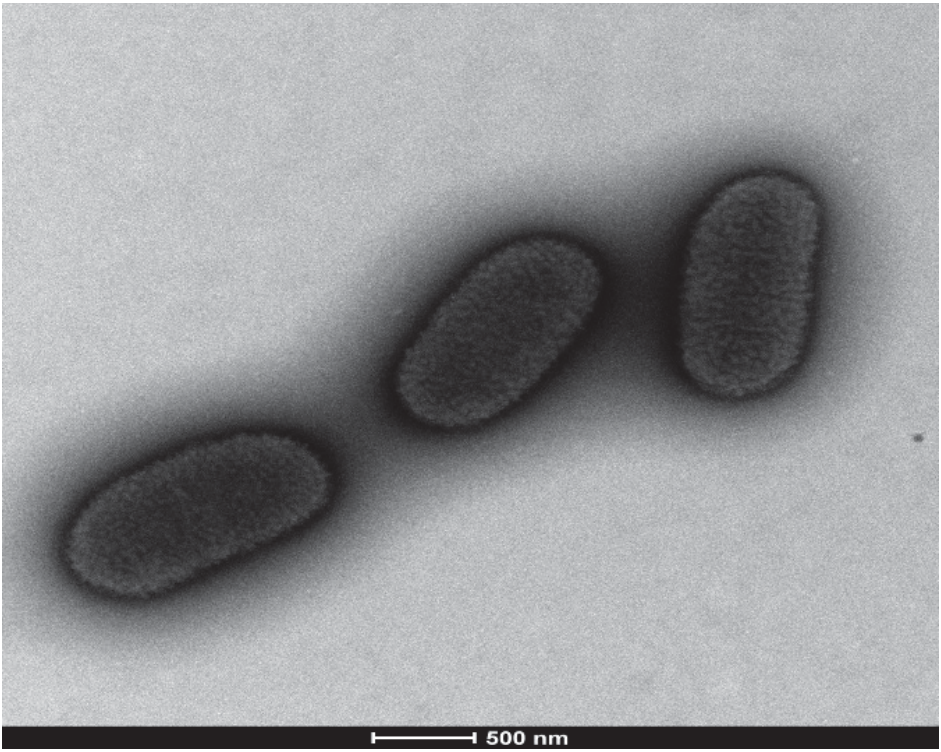
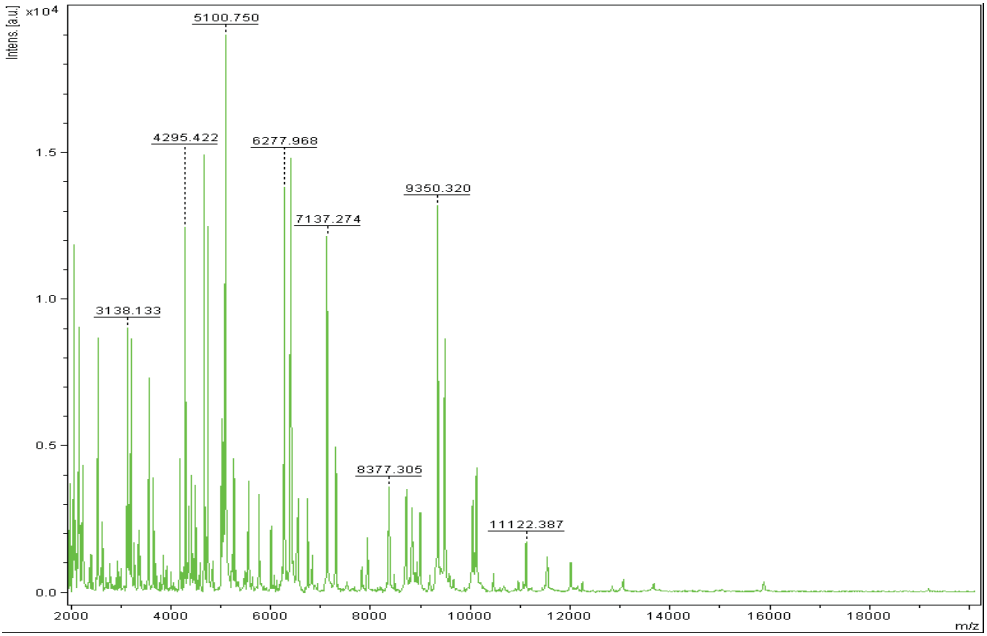
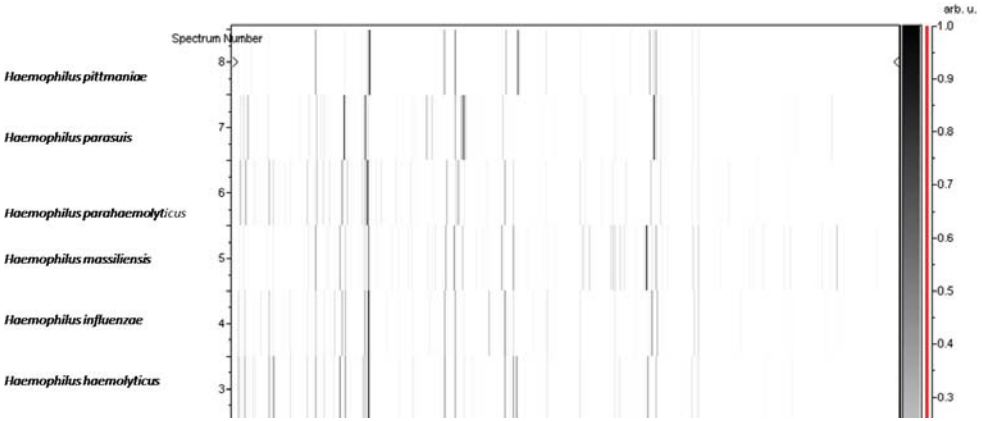


Figure 4.

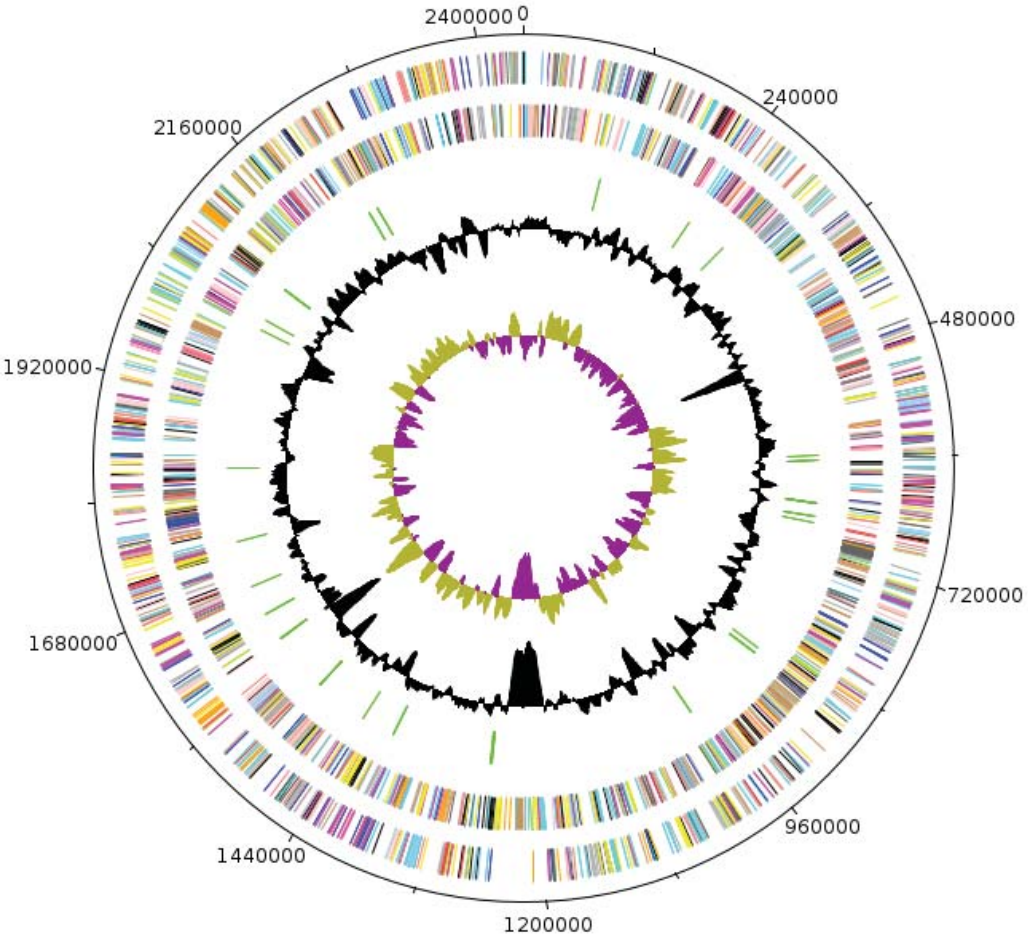

Intens.[a.u.]
x10^4
1.5
1.0
0.5
0.0
5100.750
4295.422
6277.968
9350.320
7137.274
3138.133
8377.305
11122.387
2000
4000
6000
8000
10000
12000
14000
16000
18000
m/z

509 **Figure 5.**

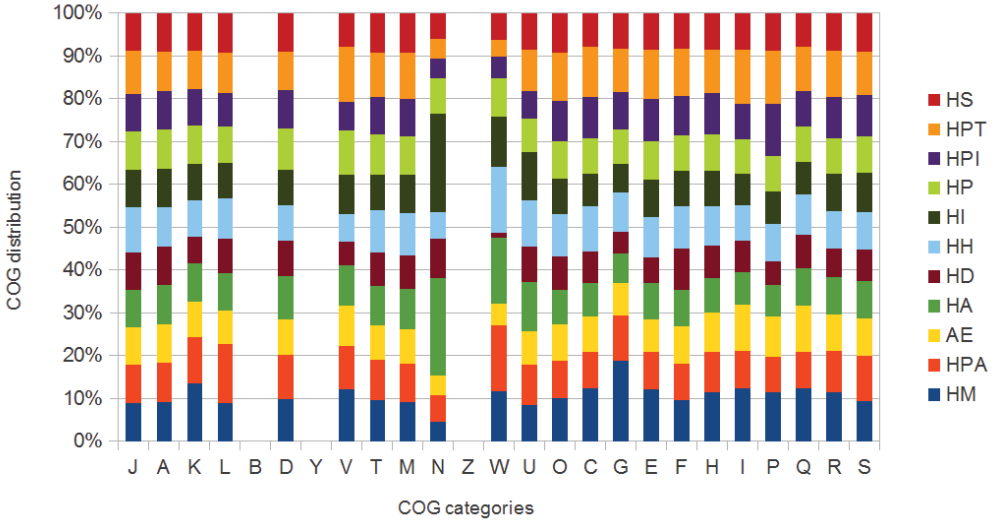


510
511

512 **Figure 6.**


2400000 0
240000
480000
720000
960000
1200000
1440000
1680000
1920000
2160000

513 **Figure 7.**
514



515
516
517

Article 11:

« Genome sequence and description of Pantoea septica strain FF5 »

Cheikh Ibrahima Lo[1], Roshan Padhmanabhan[1], Oleg Mediannikov[1, 2], Thi Tien Nguyen[1], Didier Raoult[1, 2,3], Pierre-Edouard Fournier[1] and Florence Fenollar[1,2*]

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de Médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France Tel: +33 491 385 517

[2]Campus international UCAD-IRD, Dakar, Senegal

[3]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

Publié dans:

Standards in Genomic Sciences 2015, 10:102 (14 November 2015)

http://www.standardsingenomics.com/content/10/1/103

Résumé de l'article 11: ***«Genome sequence and description of Pantoea septica strain FF5»***

La souche FF5 a été isolée de la flore cutanée d'une jeune femme Sénégalaise de 35 ans en bonne santé. Cette souche a été identifiée comme étant *Pantoea septica* par spectrométrie de masse MALDI-TOF et par séquençage de l'ARN 16S ribosomal avec une similarité de 99,5% avec la souche de *Pantoea septica* LMG 5345^{T}. Cette souche de *P. septica* est Gram négative, aérobie, mobile, et en forme de bâtonnet. Actuellement, 17 génomes ont été séquencés dans le genre *Pantoea* mais aucune de *Pantoea septica*. Nous avons comparé les propriétés génomiques de la souche FF5 à celles des autres espèces du genre *Pantoea*. Le génome de cette souche est composé de 4 548 444 paires de bases (1 chromosome, pas de plasmide) avec un contenu G+C de 59,1% codant pour 4 125 protéines et 68 gènes ARN (incluant 2 opérons ARNr). Nous avons aussi réalisé des analyses exhaustives montrant de nouvelles caractéristiques telles que la production de phosphatase alcaline, de phosphatase acide et de naphthol-AS-BI-phosphohydrolase.

Les séquences de l'ARN 16S ribosomique et du génome de cette souche de *P. septica* ont été déposées dans GenBank sous les numéros d'accession respectifs : HG315677 et CCAQ000000000.

Lo et al. Standards in Genomic Sciences (2015) 10:103
DOI 10.1186/s40793-015-0083-0

Standards in Genomic Sciences

SHORT GENOME REPORT Open Access

Genome sequence and description of *Pantoea septica* strain FF5

Cheikh Ibrahima Lo[1,2], Roshan Padhmanabhan[1], Oleg Mediannikov[1,2], Thi Tien Nguyen[1], Didier Raoult[1,2,3], Pierre-Edouard Fournier[1] and Florence Fenollar[1,2*]

Abstract

Strain FF5 was isolated from the skin flora of a healthy Senegalese 35-year-old woman. This strain was identified as belonging to the species *Pantoea septica* based on *rpoB* sequence identity of 99.7 % with *Pantoea septica* strain LMG 5345^T and a highest MALDI-TOF-MS score of 2.3 with *Pantoea septica*. Like *P. septica*, this FF5 strain is a Gram-negative, aerobic, motile, and rod-shaped bacterium. Currently, 17 genomes have been sequenced within the genus *Pantoea* but none for *Pantoea septica*. Herein, we compared the genomic properties of strain FF5 to those of other species within the genus *Pantoea*. The genome of this strain is 4,548,444 bp in length (1 chromosome, no plasmid) with a G + C content of 59.1 % containing 4125 protein-coding and 68 RNA genes (including 2 rRNA operons). We also performed an extensive phenotypic analysis showing new phenotypic characteristics such as the production of alkaline phosphatase, acid phosphatase and naphthol-AS-BI-phosphohydrolase.

Keywords: *Pantoea septica*, Genome, Taxonogenomics, Culturomics, Senegal

Introduction

Pantoea septica Brady et al. 2010 was first isolated from a human stool sample in New Jersey USA [1]. *Pantoea septica* strain FF5 (= CSUR P3024 = DSM 27843) was cultivated from the skin of a healthy Senegalese woman [2]. To date, the genus *Pantoea* consists of 22 species and 2 subspecies [3, 4] and no genome had been described for *Pantoea septica* when this paper was written. *Pantoea* species have been isolated mostly from the environment, particularly from plants, seeds and vegetables, several being phytopathogenic [5]. Some species such as *P. agglomerans, P. septica* and *P. eucrina* are also frequently isolated from humans in whom they can cause opportunistic infections [1–6].

We provide here a summary classification and a set of features for *Pantoea septica* strain FF5, together with the description of the complete genomic sequence and annotation.

Organism information

Classification and features

A skin sample was collected with a swab from a healthy Senegalese volunteer living in Dielmo (a rural village in the Guinean-Sudanian area in Senegal) in December 2012 (Table 1). This 35-year-old woman was included in a research project that was approved by the Ministry of Health of Senegal, the assembled village population and the National Ethics Committee of Senegal (CNERS, agreement numbers 09–022), as published elsewhere [7]. Strain FF5 (Table 1) was isolated by aerobic cultivation on 5 % sheep blood-enriched Columbia agar (BioMérieux, Marcy l'Etoile, France). As the 16S rRNA gene sequence cannot be used as a means of identifying *Pantoea* species, a comparative *rpoB* nucleotide sequences analysis between strain FF5 and other *Pantoea* species was performed. Strain FF5 exhibited a 99.7 % sequence identity with *P. septica*, its phylogenetically closest validly published *Pantoea* species (Fig. 1) [8]. This strain is motile and its cells grown on agar are Gram-negative rods (and have a mean diameter of 0.79-1.06 μm and a mean length of 1.25-2.04 μm).

* Correspondence: florence.fenollar@univ-amu.fr
[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
[2]Campus international UCAD-IRD, Dakar, Senegal
Full list of author information is available at the end of the article

Table 1 Classification and general features of *Pantoea septica* strain FF5 according to the MIGS recommendations [12]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain: *Bacteria*	TAS [24]
		Phylum: *Proteobacteria*	TAS [25, 26]
		Class: *Gammaproteobacteria*	TAS [26, 27]
		Order: *Enterobacteriales*	TAS [28]
		Family: *Enterobacteriaceae*	TAS [4, 28, 29]
		Genus: *Pantoea*	TAS [1]
		Species: *Pantoea septica*	IDA
		Strain: FF5	IDA
	Gram stain	Negative	IDA
	Cell shape	Rods	IDA
	Motility	Motile	IDA
	Sporulation	Non-spore forming	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37–45 °C	IDA
	pH range; Optimum	6.2–7.5; 6.8	
	Carbon source	Unknown	
MIGS-6	Habitat	Human skin	IDA
MIGS-6.3	Salinity	Growth in BHI medium + 5 % NaCl	IDA
MIGS-22	Oxygen requirement	Aerobic	IDA
MIGS-15	Biotic relationship	Free-living	IDA
MIGS-14	Pathogenicity	Unknown	
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection time	December 2012	IDA
MIGS-4.1	Latitude	13.7167	IDA
MIGS-4.1	Longitude	−16.4167	IDA
MIGS-4.4	Altitude	45 m above sea level	IDA

[a]Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [30]

Strain FF5 was catalase-positive but oxidase-negative. Using the API 20E system (BioMérieux), positive reactions were detected for β-galactosidase, citrate, tryptophan deaminase, mannitol, inositol, rhamnose, saccharose, melibiose, arabinose and sorbitol. Negative reactions were noted for arginine dehydrolase, lysine decarboxylase, hydrogen sulfide (H_2S), urease, indole and amygdalin. Using API 50 CH (BioMérieux), positive reactions were observed for glycerol, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, D-maltose, D-trehalose, D-lyxose and D-fucose. Negative reactions were observed for erytritol, L-xylose, D-adonitol, methyl β-D-xylopyranoside, L-sorbose, dulcitol, methyl α-D-mannopyranoside, methyl α-D-glucopyranoside, arbutine, salicin, D-cellobiose, inulin, D-melezitose, starch, potassium gluconate, glycogen and 5-keto-D-gluconate. Using API ZYM, positive reactions were observed for alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, acid phosphatase, naphthol-AS-BI-phosphohydrolase and . Negative reactions were observed for valine arylamidase, trypsin, α-chrymotrypsin, α-galactosidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Strain FF5 is susceptible to ceftriaxone, imipenem, gentamicin and ciprofloxacin but resistant to penicillin, amoxicillin, ticarcillin, amoxicillin-clavulanic acid, trimethoprim-sulfamethoxazole, colistin and vancomycin. Thus, the phenotypic characteristics of this strain support the claim that it belongs to *Pantoea septica*.

Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry protein analysis was performed using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany), as previously reported [9]. The scores previously established by Bruker Daltonics, used to validate or invalidate identification compared to the instrument database, were applied. Briefly, a score ≥ 2 for a species with a

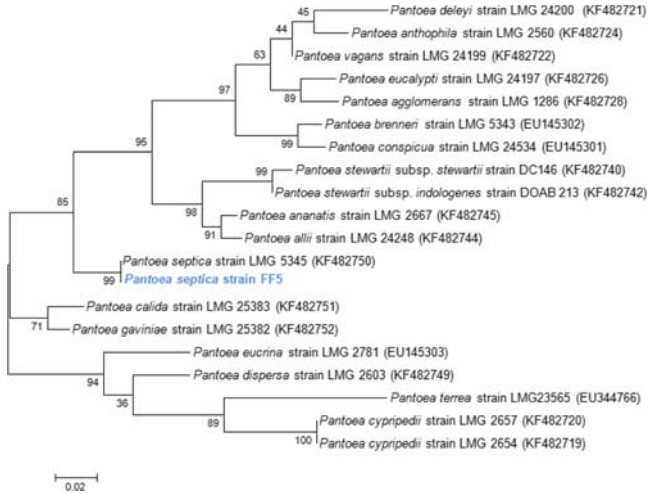


Fig. 1 Phylogenetic tree showing the position of *Pantoea septica* strain FF5 relative to other strains within the genus *Pantoea*. The *rpoB* sequences were aligned using MUSCLE [31], and the phylogenetic tree was inferred using the Maximum Likelihood method with Kimura 2-parameter model from MEGA software. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. The scale bar represents a rate of substitution per site of 0.02

validly published name provided allows the identification at the species level; a score ≥ 1.7 and < 2 allows the identification at the genus level; and a score < 1.7 does not allow any identification. Twelve distinct deposits of strain FF5 were made from 12 isolated colonies. Each smear was overlaid with 2 μL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) and dried for 5 min, as previously reported [9, 10]. The spectra from the 12 different colonies were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against the spectra of 6252 bacterial spectra. Spectra were compared with the Bruker database that contained spectra from the ten validly named *Pantoea* species. The spectra obtained were similar to those of *P. septica*. A score of 2.3 was obtained for strain FF5 supporting the identification of *P. septica*. Its reference mass spectrum was added to our database (Fig. 2).

Genome sequencing information

Genome project history

Pantoea septica strain FF5 was selected for sequencing because no genome of *P. septica* has previously been described. Besides, this strain is part of a study aiming to characterize the skin flora of healthy Senegalese people. It is the 17th genome of *Pantoea* species to be sequenced and the first genome within *P. septica*. The GenBank accession number is CCAQ000000000 and it consists of 4 scaffolds and 37 contigs. Table 2 shows the project information and its association with MIGS version 2.0 compliance [11]. Associated MIGS records are detailed in Additional file 1: Table S1.

Growth conditions and genomic DNA preparation

Pantoea septica strain FF5 (= CSUR P3024 = DSM 27843) was grown aerobically on 5 % sheep blood-enriched Columbia agar (bioMérieux) at 37 °C. Bacteria grown on four Petri dishes were resuspended in 5 × 100 μL of TE buffer; 150 μL of this suspension was diluted in 350 μL 10X TE buffer, 25 μL proteinase K and 50 μL sodium dodecyl sulfate for lysis treatment. This preparation was incubated overnight at 56 °C. DNA was purified using 3 successive phenol-chloroform extractions and ethanol precipitation at −20 °C of at least two hours each. Following centrifugation, the DNA was suspended in 65 μL EB buffer. Genomic DNA concentration was measured at 46.06 ng/μL using the Qubit assay with the high-sensitivity kit (Life technologies, Carlsbad, CA, USA).

Genome sequencing and assembly

The genomic DNA of *Pantoea septica* was sequenced using MiSeq Technology (Illumina Inc, San Diego, CA, USA) with the 2 applications: paired-end and mate-pair.

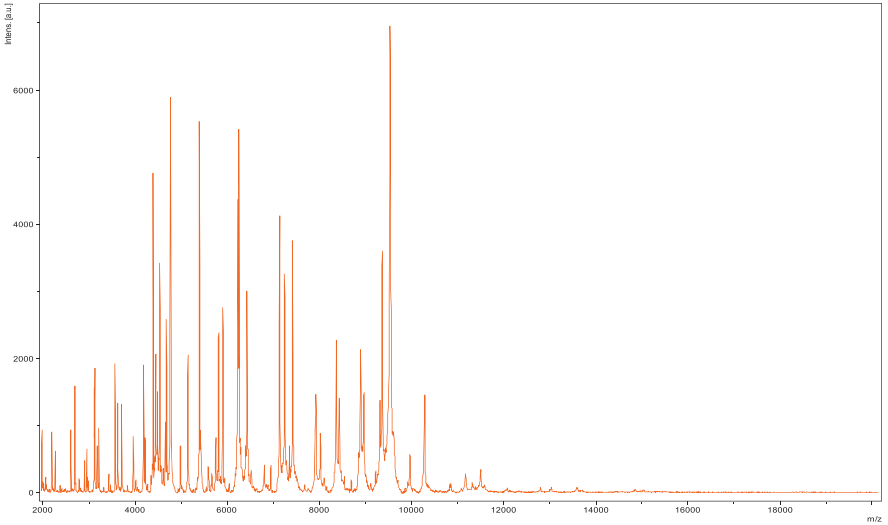


Fig. 2 Reference mass spectrum from *Pantoea septica* strain FF5. Spectra from 12 individual colonies were analyzed and a reference spectrum was generated

The paired-end and mate-pair strategies were barcoded in order to be mixed respectively with 10 other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and 11 other projects with the Nextera Mate-Pair sample prep kit (Illumina).

Genomic DNA was diluted to 1 ng/μL to prepare the paired-end library. The "tagmentation" step fragmented and tagged the DNA with an optimal size distribution of 2.25 kb. Limited cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter Inc, Fullerton, CA, USA), the libraries were normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single-strand library was loaded onto the reagent cartridge, then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in single 39-h run in 2x250-bp. Total information of 5.91 GB was obtained from a 654 K/mm2 cluster density with a cluster passing quality control filters of 93.7 % (12,204,000 clusters). Within this run, the index representation for *P. septica* was determined to be 2.25 %. So *P. septica* has 257,400 reads filtered according to the read qualities.

Table 2 Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Paired-end and mate-pair 9-kb library
MIGS-29	Sequencing platforms	MiSeq
MIGS-31.2	Fold coverage	26×
MIGS-30	Assemblers	Newbler version 2.5.3
MIGS-32	Gene calling method	Prodigal
	Locus Tag	Not indicated
	Genbank ID	CCAQ000000000
	Genbank Date of Release	March 18, 2014
	GOLD ID	Gp0100998
	BioProject ID	PRJEB4277
MIGS-13	Source material identifier	DSM 27843
	Project relevance	Study of human skin flora

The mate pair library was prepared with 1 μg of genomic DNA using the Nextera mate-pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The fragmentation profile was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1.5 kb up to 14 kb with an optimal size of 9 kb. No size selection was performed and 600 ng of tagmented fragments were circularized. The

circularized DNA was mechanically sheared into small fragments on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High-Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA). The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution to 10 pM, the pool of libraries was loaded onto the reagent cartridge, then onto the instrument along with the flow cell. Automated cluster generation and sequencing were performed in a single 39-h run in a 2x250-bp.

An overall quantity of 3.2 GB was obtained from a 690 K/mm2 cluster density with a cluster passing quality control filters of 95.4 % (13,264,000 clusters). The index representation for *P. septica* was determined to be 7.26 % within this run. *P. septica* has a total of 918,753 reads filtered according to the read qualities.

Genome annotation

Open Reading Frames prediction was performed using Prodigal [12] with default parameters. We removed the predicted ORFs if they spanned a sequencing gap region. Functional assessment of protein sequences was performed by comparing them with sequences in the GenBank [13] and Clusters of Orthologous Groups (COG) databases using BLASTP. tRNAs, rRNAs, signal peptides and transmembrane helices were identified using tRNAscan-SE 1.21 [14], RNAmmer [15], SignalP [16] and TMHMM [17] respectively. Artemis [18] was used for data management whereas DNA Plotter [19] was used for visualization of genomic features. In-house perl and bash scripts were used to automate these routine tasks. ORFans were sequences with no homology in a given database i.e. in a non-redundant (nr) or identified if their BLASTP E-value was lower than 1e-03 for alignment lengths greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E-value of 1e-05. PHAST was used to identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids [20].

To estimate the nucleotide sequence similarity at the genome level between *P. septica* and another 7 members of the genus of *Pantoea* and 4 members of the genus *Enterobacter*, we determined the AGIOS parameter as follows: orthologous proteins were detected using the Proteinortho software (with the parameters following: E-value 1e-5, 30 % identity, 50 % coverage and algebraic connectivity of 50 %) [21] and genomes compared two by two. After fetching the corresponding nucleotide sequences of orthologous proteins for each pair of genomes, we determined the mean percentage of nucleotide sequence identity using the Needleman-Wunsch global alignment algorithm. The script created to calculate AGIOS values was named MAGi (Marseille Average genomic identity) and is written in perl and bioperl modules. GGDC analysis was also performed using the GGDC web server as previously reported [22].

Genome properties

The genome of *P. septica* strain FF5 is 4,548,444 bp long (1 chromosome, no plasmid) with a 59.1 % G + C content (Fig. 3). Of the 4193 predicted genes, 4125 were protein-coding genes and 68 were RNAs. A total of 3040 genes (72.50 %) were assigned a putative function. A total of 522 genes were annotated as hypothetical proteins. The properties and statistics of the genome are presented in Table 3. The distribution of genes into COG functional categories is presented in Table 4. A total of 214 were identified as ORFans (5.18 %).

Insights from genome sequence

Here, we compared 11 genome sequences including *Pantoea ananatis* strain LMG 20103, *P. vagans* strain C9-1, *P. ananatis* strain LMG 5342, *P. ananatis* strain AJ13355, *P. ananatis* strain PA13, *P. agglomerans* strain 299R, *P. stewartii* subsp. *stewartii* strain DC283, *Enterobacter cloacae* subsp. *dissolvens* strain SDM, *E. aerogenes* strain EA1509E, *E. asburiae* strain LF7a and *E. cloacae* strain EcWSU1 (Table 5).

Table 5 shows a comparison of genome size, G + C content, coding-density and number of proteins for these genomes.

The G + C content (59.1 %) of *P. septica* strain FF5 differed by more than 1 % from all other compared species within the genus *Pantoea* [*P. vagans* strain C9-1 (55.55), *P. ananatis* strains LMG 5342, AJ13355 and PA13 (53.45, 53.76, and 53.66, respectively), *P. agglomerans* strain 299R (54.3), *P. stewartii* subsp. *stewartii* strain DC283 (53.8)].

According to the previous demonstration that the G + C content deviation is at most 1 % within species, these values confirm the classification of strain FF5 in a distinct species [23].

Orthologous gene comparison of *P. septica* strain FF5 with other closely related species are summarized in Table 6. Intraspecies values ranged from 99.06 to 99.33 % for *P. ananatis* (Table 7). Interspecies AGIOS values ranged from 77.46 to 84.94 % within the *Pantoea* genus, and from 71.27 to 72.57 % between *Pantoea* and *Enterobacter* species (Table 7). When compared to other species, *P. septica* exhibited AGIOS values ranging from 77.7 to 80.5 with *Pantoea* species and from 72.38 to 73.26 with *Enterobacter* species (Table 7).

Conclusions

We describe the genome of *Pantoea septica* strain FF5. This is the first reported genome of *P. septica*. We also

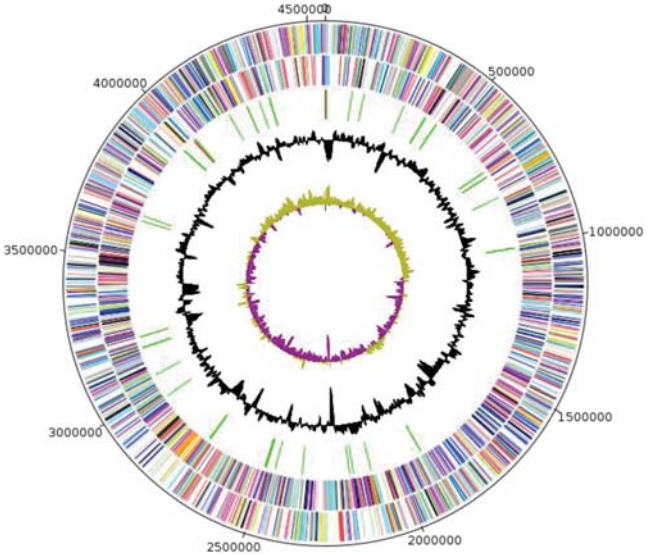


Fig. 3 Graphical circular map of the chromosome of *P. septica* strain FF5. From the outside in the two outer circles show open reading frames oriented in the forward (colored by COG categories) and reverse (colored by COG categories) directions, respectively. The third circle marks the rRNA gene operon (red) and tRNA genes (green). The fourth circle shows the G + C% content plot. The innermost circle shows GC skew, with purple and olive indicating negative and positive values, respectively

Table 3 Nucleotide content and gene count levels of the genome

Attribute	Value	% of total[a]
Genome size (bp)	4,548,444	
DNA coding (bp)	3,981,573	87.54
DNA G + C (bp)	2,687,917	59.1
DNA scaffolds	4	-
Total genes	4,193	100.00
Protein-coding genes	4,125	98.37
RNA genes	68	1.50
Pseudo genes	22	0.53
Genes in internal clusters	N/D[b]	-
Genes with function prediction	3,040	72.50
Genes assigned to COGs	3,562	84.97
Genes with Pfam domains	134	3.24
Genes with peptide signals	214	5.18
Genes with transmembrane helices	1,026	24.87
ORFan genes	532	12.89
CRISPR repeats	3	

[a]The total is based on either the size of genome in base pairs or the total number of protein-coding genes in the annotated genome

Table 4 Number of genes associated with general COG functional categories

Code	Value	% age	Description
J	173	4.19	Translation, ribosomal structure and biogenesis
A	1	0.02	RNA processing and modification
K	274	6.64	Transcription
L	118	2.86	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	33	0.80	Cell cycle control, Cell division, chromosome partitioning
V	41	0.99	Defense mechanisms
T	99	2.40	Signal transduction mechanisms
M	191	4.63	Cell wall/membrane biogenesis
N	41	0.99	Cell motility
Z	0	0.00	Cytoskeleton
U	34	0.82	Intracellular trafficking and secretion
O	113	2.73	Posttranslational modification, protein turnover, chaperones
C	186	4.50	Energy production and conversion
G	233	5.65	Carbohydrate transport and metabolism
E	309	7.49	Amino acid transport and metabolism
F	75	1.82	Nucleotide transport and metabolism
H	113	2.74	Coenzyme transport and metabolism
I	69	1.67	Lipid transport and metabolism
P	207	5.01	Inorganic ion transport and metabolism
Q	36	0.87	Secondary metabolite biosynthesis, transport and catabolism
R	363	8.80	General function prediction only
S	331	8.02	Function unknown
-	522	12.65	Not in COGs

The total is based on the total number of protein-coding genes in the annotated genome

Table 5 Comparison of *Pantoea septica* strain FF5 with other genomes of several *Pantoea* species and some *Enterobacter* species

Microorganisms used for genome comparison	Accession number	Genome size (bp)	GC%	Number of proteins
P. septica strain FF5	CCAQ000000000	4,548,444	59.10	4,125
P. ananatis strain LMG 20103	NC_013956	4,703,373	53.69	4,241
P. vagans strain C9-1	NC_014562	4,024,986	55.55	3,664
P. ananatis strain LMG 5342	NC_016816	4,605,545	53.45	4,324
P. ananatis strain AJ13355	NC_017531	4,555,536	53.76	3,760
P. ananatis strain PA13	NC_017554	4,586,378	53.66	4,130
P. agglomerans strain 299R	ANKX00000000	4,581,483	54.30	4,157
P. stewartii subsp. *stewartii* strain DC283	AHIE000000000	5,233,214	53.80	4,903
E. cloacae subsp. *dissolvens* strain SDM	NC_018079	4,968,248	55.06	4,542
E. aerogenes strain EA1509E	NC_020181	5,419,609	54.98	5,260
E. asburiae strain LF7a	NC_015968	4,812,833	53.85	4,409
E. cloacae strain EcWSU1	NC_016514	4,734,438	54.61	4,534

Table 6 Orthologous gene comparison of *Pantoea septica* strain FF5 with other closely related species

	P. septica	*P. agglomerans*	*P. stewartii*	*P. ananatis* LMG20103	*P. vagans* C9	*P. ananatis* LMG5342	*P. ananatis* AJ13355	*P. ananatis* PA13	*E. cloacae* SDM	*E. aerogenes* EA1509E	*E. asburiae* ELF7a	*E. cloacae* EcWSU1
P. septica	**4,125**											
P. agglomerans	2,948	**4,157**										
P. stewartii	2,677	2,581	**4,903**									
P. ananatis LMG20103	2,993	2,953	3,024	**4,241**								
P. vagans-C9	2,928	2,889	2,576	2,889	**3,664**							
P. ananatis-LMG5342	2,868	2,792	2,917	3,527	2,852	**4,324**						
P. ananatis AJ13355	2,778	2,698	2,775	3,372	2,752	3,413	**3,760**					
P. ananatis PA13	2,876	2,801	2,960	3,560	2,883	3,648	3,402	**4,130**				
E. cloacae SDM	2,736	2,536	2,400	2,688	2,535	2,586	2,549	2,585	**4,542**			
E. aerogenes EA1509E	2,688	2,495	2,400	2,672	2,528	2,617	2,570	2,612	3,282	**5,260**		
E. asburiae ELF7a	2,634	2,471	2,393	2,634	2,502	2,577	2,542	2,588	3,650	3,249	**4,409**	
E. cloacae EcWSU1	2,674	2,526	2,387	2,664	2,529	3,456	2,558	2,520	2,572	3,457	3,105	**4,534**

Bold numbers indicate the number of genes from each genome

Table 7 dDDH values (upper right) and AGIOS values (lower left) obtained by comparison of all studied genomes

	P. septica	*P. ananatis* LMG 20103	*P. vagans* C9-1	*P. ananatis* LMG 5342	*P. ananatis* AJ13355	*P. ananatis* A13	*P. agglomerans* 299R	*P. stewartii* DC283	*E. cloacae* SDM	*E. aerogenes* EA1509E	*E. asburiae* LF7a	*E. cloacae*EcWSU1
P. septica		0.2038	0.1913	0.2041	0.2036	0.2033	0.1966	0.203	0.2152	0.2091	0.2182	0.2174
P. ananatis LMG 20103	77.7		0.1916	0.0084	0.0077	0.0089	0.1955	0.1534	0.2127	0.2026	0.2158	0.2151
P. vagans C9-1	80.5	79.69		0.1907	0.1908	0.1907	0.0935	0.191	0.214	0.2126	0.2121	0.2125
P. ananatis LMG 5342	78.11	99.14	79.85		0.0094	0.0099	0.1956	0.1519	0.2136	0.2027	0.2177	0.2133
P. ananatis AJ13355	78.17	99.33	79.96	99.33		0.009	0.1959	0.1523	0.2144	0.2032	0.2176	0.2131
*P. ananatis*PA13	78.06	99.07	79.81	99.07	99.11		0.196	0.1519	0.2145	0.2032	0.216	0.2139
P. agglomerans 299R	79.12	78.75	91.2	79.14	79.22	78.06		0.1973	0.2197	0.2207	0.2208	0.222
P. stewartii DC283	78.01	84.54	79.79	84.73	84.94	84.6	78.99		0.2136	0.2025	0.2183	0.2134
E. cloacae SDM	72.79	71.6	72.57	71.64	71.79	71.68	71.92	71.22		0.1917	0.1379	0.1194
E. aerogenes EA1509E	73.26	71.48	72.37	71.44	71.58	71.41	71.76	71.53	78.09		0.1955	0.195
E. asburiae LF7a	72.38	71.38	72.22	71.34	71.44	71.27	71.77	71.52	85.85	77.73		0.1394
E. cloacae EcWSU1	72.68	71.52	72.38	85.73	71.59	71.74	71.76	71.67	71.53	87.91	78.38	

report phenotypic and phylogenetic characteristics of strain FF5. *P. septica* strain FF5 was isolated from the skin flora of a 35-year-old healthy Senegalese woman. The *P. septica* strain FF5 genome sequences are deposited in GenBank under accession number CCAQ000000000.

Additional file

Additional file 1: Table S1. Associated MIGS record. (DOC 70 kb)

Abbreviations

DSM: Deutsche Sammlung von Mikroorganismen; CSUR: Collection de Souches de l'Unité des Rickettsies; MALDI: Matrix Assisted Laser Desorption Ionization; AGIOS: Average Genomic Identity of Orthologous Gene Sequences; GGDC: Genome-to-Genome Distance Calculator; dDDH: Digital DNA-DNA hybridization; MIGS: Minimum Information about a Genome Sequence.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CIL performed the phenotypic characterization of the bacterium and drafted the manuscript. RP performed the genomic analyses and drafted the manuscript. OM participated in its design and helped to draft the manuscript. TTN performed the genomic sequencing and helped to draft the manuscript. DR conceived the study and helped to draft the manuscript. PEF and FF conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Dr Carine Couderc for her help in performing the MALDI-TOF analysis. This study was funded by the Méditerranée Infection Foundation.

Author details

[1]Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France. [2]Campus international UCAD-IRD, Dakar, Senegal. [3]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia.

Received: 3 July 2014 Accepted: 19 October 2015
Published online: 14 November 2015

References

1. Brady CL, Cleenwerck I, Venter SN, Engelbeen K, De Vos P, Coutinho TA. Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of *Pectobacterium cypripedii* (Hori 1911) Brenner et al. 1973 emend. Hauben et al. 1998 to the genus as *Pantoea cypripedii* comb. nov. Int J Syst Evol Microbiol. 2010;60:2430–40. doi:10.1099/ijs.0.017301-0.
2. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012;18:1185–93. doi:10.1111/1469-0691.12023.
3. Parte AC. LPSN-list of prokaryotic names with standing in nomenclature. Nucleic Acids Res. 2014;42:D613–6. doi:10.1093/nar/gkt1111.
4. Euzéby JP. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. Int J Syst Bacteriol. 1997; 47:590–2.
5. Mergaert J, Verdonck L, Kersters K. Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the Genus *Pantoea* emend. as *Pantoea ananas* (Serrano1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and Description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. Int J Syst Bacteriol. 1993;43:162–73. doi:10.1099/00207713-43-1-162.
6. Liberto MC, Matera G, Puccio R, Lo Russo T, Colosimo E, Focà E. Six cases of sepsis caused by *Pantoea agglomerans* in a teaching hospital. New Microbiol. 2009;32:119–23.
7. Trape JF, Tall A, Diagne N, Ndiath O, Ly AB, Faye J, et al. Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. Lancet Infect Dis. 2011;11:925–32. doi:10.1016/S1473-3099(11)70194-3.
8. Brady CL, Venter SN, Cleenwerck I, Engelbeen K, Vancanneyt M, Swings J, et al. *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. Int J Syst Evol Microbiol. 2009;59:2339–45.
9. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009;49:543–51. doi:10.1086/600885.
10. Fall B, Lo CI, Samb-Ba B, Perrot N, Diawara S, Gueye MW, et al. The ongoing revolution of maldi-tof mass spectrometry for microbiology reaches tropical Africa. Am J Trop Med Hyg. 2015;92:641–7.
11. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:541–7. doi:10.1038/nbt1360.
12. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
13. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2012;40:48–53. doi:10.1093/nar/gkr1202.
14. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
15. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8. doi:10.1093/nar/gkm160.
16. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95. doi:10.1016/j.jmb.2004.05.028.
17. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80. doi:10.1006/jmbi.2000.4315.
18. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics. 2000;16:944–5. doi:10.1093/bioinformatics/16.10.944.
19. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics. 2009;25:119–20. doi:10.1093/bioinformatics/btn578.
20. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39:347–52. doi:10.1093/nar/gkr485.
21. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;12:124. doi:10.1186/1471-2105-12-124.
22. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14:60.
23. Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G + C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol. 2014;64:352–6.
24. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eukarya*. Proc Natl Acad Sci U S A. 1990;87:4576–9. doi:10.1073/pnas.87.12.4576.
25. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 2, Part B. New York: Springer; 2005. p. 1.
26. Euzéby J. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. Int J Syst Evol Microbiol. 2005;55:983–5.
27. Garrity GM, Bell JA, Class LT, III. *Gammaproteobacteria* class. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. Bergey's Manual of Systematic Bacteriology, vol. 2. 2nd ed. New York: Springer; 2005. p. 1.
28. Garrity GM, Holt JG. Taxonomic Outline of the *Archaea* and *Bacteria*. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's Manual of

Systematic Bacteriology, vol. 1. 2nd ed. New York: Springer; 2001. p. 155–66.
29. Lapage SP. Proposal of *Enterobacteraceae* nom. nov. as a substitute for the illegitimate but conserved name *Enterobacteriaceae* Rahn 1937. Request for an opinion. Int J Syst Bacteriol. 1979;29:265–6.
30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9. doi:10.1038/75556.
31. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–7.

Conclusion et Perspectives générales de la thèse

L'installation et l'utilisation du MALDI-TOF en Afrique a été une grande réussite. Nos différents travaux effectués ont montré la faisabilité et la robustesse de cet appareil dans l'identification de routine des microorganismes. Il a été aussi démontré que le MALDI-TOF est une technologie qui permet une identification rapide des bactéries et champignons ainsi que d'autres organismes pour un faible coût. Il faut reconnaitre aussi que l'apport du MALDI-TOF a été significatif pour la prise en charge des patients, parce que participant considérablement à la réduction du délai de rendu des résultats de diagnostic. La rapidité et l'efficacité de l'instrument a précipité son utilisation comme outil de première ligne pour le diagnostic de routine à l'Hôpital Principal de Dakar. D'ailleurs, certaines méthodes conventionnelles telles que les galeries API, qui sont des méthodes longues et fastidieuses, ont été abandonnées au profit de l'identification par la spectrométrie de masse. Nous avons également prouvé que MALDI-TOF est un outil puissant pour identifier les espèces bactériennes rarement impliquées dans les maladies infectieuses humaines. Nous pouvons ainsi dire que le MALDI-TOF est un outil qui peut contribuer à la lutte contre les maladies infectieuses et tropicales émergentes qui sévissent encore en Afrique tropicale.

L’acquisition d'un spectromètre de masse de type MALDI-TOF, qui coûte plus de cent milles euros, pour un laboratoire de microbiologie en Afrique, sera plus facilitée si on sollicite l'aide de partenaires au développement comme les fondations et les organisations non gouvernementales. En revanche, il sera impératif de résoudre certaines contraintes techniques liées à la maintenance et à l'électricité qui participeront à la pérennité de l'appareil. Nous avons aussi démontré que le MALDI-TOF permet une identification à haut débit des microorganismes et a ainsi contribué grandement à élargir le répertoire des bactéries isolées en Afrique. En effet, en ce qui nous concerne, nous avons décrit sur la base des résultats du MALDI-TOF, au total sept nouvelles bactéries toutes isolées chez l'homme pour la première fois. Ce qui prouve encore une fois son intérêt dans le développement de la microbiologie clinique.

Dans l'avenir, il serait souhaitable d'augmenter le nombre d'instruments pour faire bénéficier des avantages de l'utilisation du MALDI-TOF à de nombreux centres de santé en Afrique. En outre, nous suggérons aux villes ayant de nombreuses structures de santé et de recherche d’utiliser une plateforme commune MALDI-TOF; ce qui permet une meilleure gestion de l’appareil et d’économiser de l’argent en réduisant le nombre de spécialistes intervenants.

Références

1. Seng P, Raoult D. Application de la spectrométrie de masse MALDI-TOF en microbiologie clinique. Marseille, France; 2013. 107 p.

2. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis Off Publ Infect Dis Soc Am. 15 août 2009;49(4):543‑51.

3. Anhalt JP, Fenselau C. Identification of bacteria using mass spectrometry. Anal Chem. 1 févr 1975;47(2):219‑25.

4. Carbonnelle E, Raskine L. Applications de la spectrométrie de masse de type MALDI-TOF en microbiologie. Bio Trib Mag. 2011;39(1):35‑42.

5. Seng P, Rolain J-M, Fournier PE, La Scola B, Drancourt M, Raoult D. MALDI-TOF-mass spectrometry applications in clinical microbiology. Future Microbiol. nov 2010;5(11):1733‑54.

6. Kliem M, Sauer S. The essence on mass spectrometry based microbial diagnostics. Curr Opin Microbiol. juin 2012;15(3):397‑402.

7. Mellmann A, Bimet F, Bizet C, Borovskaya AD, Drake RR, Eigner U, et al. High interlaboratory reproducibility of matrix-assisted laser desorption ionization-time of flight mass spectrometry-based species identification of nonfermenting bacteria. J Clin Microbiol. nov 2009;47(11):3732‑4.

8. Santos C, Paterson RRM, Venâncio A, Lima N. Filamentous fungal characterizations by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Appl Microbiol. févr 2010;108(2):375‑85.

9. Stevenson LG, Drake SK, Shea YR, Zelazny AM, Murray PR. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of clinically important yeast species. J Clin Microbiol. oct 2010;48(10):3482‑6.

10. van Veen SQ, Claas ECJ, Kuijper EJ. High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight mass spectrometry in conventional medical microbiology laboratories. J Clin Microbiol. mars 2010;48(3):900‑7.

11. Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. J Phys Chem. 1 sept 1984;88(20):4451‑9.

12. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem. 1 oct 1988;60(20):2299‑301.

13. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1 août 1988;2(8):151‑3.

14. Karas M, Bachmann D, Hillenkamp F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. Anal Chem. 1 déc 1985;57(14):2935‑9.

15. Schey K, Cooks RG, Grix R, Wöllnik H. A tandem time-of-flight mass spectrometer for surface-induced dissociation. Int J Mass Spectrom Ion Process. 1 juill 1987;77(1):49‑61.

16. Tanaka K. The origin of macromolecule ionization by laser irradiation (Nobel lecture). Angew Chem Int Ed Engl. 25 août 2003;42(33):3860‑70.

17. Riat A, Cherkaoui A, Emonet S, Greub G, Schrenzel J.Quels bénéfices pour les cliniciens de la mise en place du MALDI-TOF/MS dans le laboratoire de bactériologie ? Rev Médicale Suisse. 12 nov 2014;450(39):2149‑54.

18. Asakawa D, Bernevic B, Dubois D. Applications of MALDI-TOF spectroscopy. Cai ZA, Liu S, éditeurs. Heidelberg, Allemagne, Etats-Unis, Pays-Bas; 2013. vii+215 p.

19. Courcol R. Quelles utilisations de la spectrométrie de masse de type MALDI-TOF en microbiologie médicale ? Rev Francoph Lab. nov 2009;2009(416):61‑4.

20. Suarez S, Nassif X, Ferroni A. [Applications of MALDI-TOF technology in clinical microbiology.]. Pathol Biol (Paris). févr 2015;63(1):43‑52.

21. Alispahic M, Hummel K, Jandreski-Cvetkovic D, Nöbauer K, Razzazi-Fazeli E, Hess M, et al. Species-specific identification and differentiation of Arcobacter, Helicobacter and Campylobacter by full-spectral matrix-associated laser desorption/ionization time of flight mass spectrometry analysis. J Med Microbiol. mars 2010;59(Pt 3):295‑301.

22. Winkler MA, Uher J, Cepa S. Direct analysis and identification of Helicobacter and Campylobacter species by MALDI-TOF mass spectrometry. Anal Chem. 15 août 1999;71(16):3416‑9.

23. Tindall BJ. Updating Rule 15 of the International Code of Nomenclature of Bacteria. Int J Syst Evol Microbiol. 7 mai 2015;ijs.0.000310.

24. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored « rare biosphere ». Proc Natl Acad Sci U S A. 8 août 2006;103(32):12115‑20.

25. Carroll K, Weinstein M. Manual and automated systems for detection and identification of microorganisms. In American Society for Microbiology WD- (ed.), Manual of Clinical Microbiology, 9th Edition ed. In 2007.

26. Bessède E, Angla-gre M, Delagarde Y, Sep Hieng S, Ménard A, Mégraud F. Matrix-assisted laser-desorption/ionization BIOTYPER:

experience in the routine of a University hospital. Clin Microbiol Infect. 1 avr 2011;17(4):533‑8.

27. Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen K-Y. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. oct 2008;14(10):908‑34.

28. Schlaberg R, Simmon KE, Fisher MA. A systematic approach for discovering novel, clinically relevant bacteria. Emerg Infect Dis. mars 2012;18(3):422‑30.

29. Greub G. MALDI-TOF mass spectrometry: the quantum leap. Clin Microbiol Infect. nov 2010;16(11):1603.

30. Moon H-W, Lee SH, Chung H-S, Lee M, Lee K. Performance of the Vitek MS matrix-assisted laser desorption ionization time-of-flight mass spectrometry system for identification of Gram-positive cocci routinely isolated in clinical microbiology laboratories. J Med Microbiol. sept 2013;62(Pt 9):1301‑6.

31. Manji R, Bythrow M, Branda JA, Burnham C -a. D, Ferraro MJ, Garner OB, et al. Multi-center evaluation of the VITEK® MS system for mass spectrometric identification of non-Enterobacteriaceae Gram-negative bacilli. Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol. mars 2014;33(3):337‑46.

32. Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev. juill 2013;26(3):547‑603.

33. Garner O, Mochon A, Branda J, Burnham C-A, Bythrow M, Ferraro M, et al. Multi-centre evaluation of mass spectrometric identification of anaerobic bacteria using the VITEK® MS system. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. avr 2014;20(4):335‑9.

34. Fall B, Lo CI, Samb-Ba B, Perrot N, Diawara S, Gueye MW, et al. The Ongoing Revolution of MALDI-TOF Mass Spectrometry for Microbiology Reaches Tropical Africa. Am J Trop Med Hyg. 2015;14‑0406.

35. L'aquariophilie pour des aquariums modernes : définition des espèces bactériennes [Internet]. Aquaportail.com. 2015 [cité 5 juill 2015]. Disponible sur: http://www.aquaportail.com/definition-12741-especes-bacteriennes.html#ixzz3ejWnUoYj

36. Decoster A, Lemahieu J-C. Cours de microbiologie de la Faculté Libre de Médecine de Lille [Internet]. Classification des bactéries. 2008 [cité 2 juill 2015]. Disponible sur: http://anne.decoster.free.fr/btelechar/bpoly/Taxono01.pdf

37. Lennette EH, al. Manual of Clinical Microbiology 4th Ed. American Society for Microbiology, Washington, D.C. 1985.

38. Baron EJ. Classification. In: Baron S, editor. Medical Microbiology. 4th edition.Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 3. 1996.

39. Schleifer KH. Classification of Bacteria and Archaea: Past, present and future. Syst Appl Microbiol. déc 2009;32(8):533‑42.

40. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiology Today. 2006;152‑5.

41. Renvoisé A. Applicabilité de la PCR « universelle » 16S comme outil d'identification et de détection bactérienne en laboratoire hospitalier de bactériologie [Internet]. Aix-Marseille; 2012 [cité 5 juill 2015]. Disponible sur: http://www.theses.fr/2012AIXM5016

42. Clarridge JE. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. Clin Microbiol Rev. 10 janv 2004;17(4):840‑62.

43. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods. mai 2007;69(2):330‑9.

44. Fenollar F, Roux V, Stein A, Drancourt M, Raoult D. Analysis of 525 samples to determine the usefulness of PCR amplification and sequencing of the 16S rRNA gene for diagnosis of bone and joint infections. J Clin Microbiol. mars 2006;44(3):1018‑28.

45. Drancourt M, Berger P, Raoult D. Systematic 16S rRNA gene sequencing of atypical clinical isolates identified 27 new bacterial species associated with humans. J Clin Microbiol. mai 2004;42(5):2197‑202.

46. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol. sept 2007;45(9):2761‑4.

47. Olsen GJ, Woese CR. Ribosomal RNA: a key to phylogeny. FASEB J Off Publ Fed Am Soc Exp Biol. janv 1993;7(1):113‑23.

48. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. janv 1991;173(2):697‑703.

49. Janda JM, Abbott SL. Bacterial identification for publication: when is enough enough? J Clin Microbiol. juin 2002;40(6):1887‑91.

50. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kämpfer P, Maiden MCJ, et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol. mai 2002;52(Pt 3):1043‑7.

51. Lo CI, Mishra AK, Padhmanabhan R, Samb B, Sow AG, Robert C, et al. Non-contiguous finished genome sequence and description of Clostridium dakarense sp. nov. Stand Genomic Sci. 16 oct 2013;9(1):14‑27.